UNIVERSITY OF ARIZONA
39001004869551

PRO ENCE TEACHING *at the* COLLEGE LEVEL

A. W. HURD

COLLEGE PROBLEMS SERIES

THE TEACHING OF SCIENCE AT THE COLLEGE LEVEL

EDITED BY E. M. FREEMAN
DEAN OF THE COLLEGE OF AGRICULTURE
UNIVERSITY OF MINNESOTA

VOLUME I

Q
181
H97
1928

PROBLEMS OF SCIENCE TEACHING AT THE COLLEGE LEVEL

By

ARCHER WILLIS HURD

Research Associate, Teachers College
Columbia University

MINNEAPOLIS
THE UNIVERSITY OF MINNESOTA PRESS
1929

COPYRIGHT, 1929
BY THE UNIVERSITY OF MINNESOTA
ALL RIGHTS RESERVED

FOREWORD

Few things have been more gratifying to the Committee on Educational Research than the generous response of the teachers of science in the University to the suggestion that the problems of science teaching should be studied. An adherent of the theory of the transfer of training might find in the willingness of science teachers to experiment with educational problems an illustration of his philosophy. The Chairman of this Committee would, however, call attention to the additional fact that the particular scientists who have cooperated with Dr. Hurd in the investigations here reported are men with a deep interest in the students whom they teach and with a desire to understand the effects of their teaching upon student accomplishment. As these studies have progressed, the spirit of sound educational inquiry has deepened in the individuals most active in the investigations and has spread to others in the same and in related fields.

The Minnesota studies in the teaching of science have achieved a threefold result. In the first place they have produced certain facts that throw a new light upon the results of science teaching at the college level. Secondly, they have clarified the community of interest between students of education and college teachers and have given a useful demonstration of the possibility of cooperative research in the field of joint interest. Finally, they have revealed the intricacy and involved nature of the problem of college education and have pointed with unmistakable force to the importance of extended research in these matters.

It is obvious that college problems must be studied in great detail and by all the controlled methods that may be employed for the purpose. This means a continuing program of research reaching into the years ahead; it implies further cooperative studies, but it also indicates that men must train themselves in the techniques of educational research at the college level; it further suggests the need for increased administrative support in financial and other ways.

College education in America suffers unfavorable comparison in this matter of research into its own problems when contrasted

with American industry at its best. Industrial corporations have found it commercially profitable to subsidize in generous fashion industrial laboratories, each in connection with its own specific problems. The sums devoted to this purpose seem fabulous to a college professor or administrator. In the case of one corporation, the annual budget of its laboratories reaches $19,000,000. When colleges and organizations of colleges come to view their own mission with equal seriousness, we may expect a vast increment to our knowledge about college matters and a great improvement in the procedures of higher education.

M. E. Haggerty, *Chairman*
Committee on Educational Research
University of Minnesota
April, 1929

ACKNOWLEDGMENTS

Studies of this kind are made possible only through the cooperation of many individuals. The author wishes to make grateful acknowledgment to all persons cooperating in any way, but particularly to the following:

Dean M. E. Haggerty, adviser and counselor, who has made the study possible and stimulated its pursuance at all times; Dean E. M. Freeman, Chairman of the Subcommittee of Science of the Committee of Educational Research, and other members of the subcommittee; Dean E. P. Lyon of the Medical School and Dr. F. H. Scott of the Department of Physiology; Dr. C. M. Jackson, Head of the Department of Anatomy; Dr. H. A. Erikson, Chairman of the Department of Physics, and Dr. J. W. Buchta, Dr. L. F. Miller, and Dr. A. Zeleny of the Physics Department; Dr. W. S. Miller, Dr. Earl Hudelson, and Dr. Fred Engelhardt, members of the thesis committee, who have made helpful suggestions; Mrs. Edith M. Hurd, who has given efficient aid in much of the work of preparation of the manuscript.

A. W. Hurd

University of Minnesota
April, 1929

CONTENTS

PAGE

FOREWORD v

ACKNOWLEDGMENTS vii

CONTENTS ix

LIST OF TABLES xiii

CORRELATION CHARTS xix

PART I

TEACHING PROBLEMS

INTRODUCTION 1

Scientific method, 1. Problems in teaching, 2. Committee on educational research, 4. Program of investigations, 5. Purpose of present book, 6.

CHAPTER

I. NEED FOR EXPERIMENTATION 8

Present status of experimentation, 8. Function of the laboratory, 9. Points of view, 10. Values of certain features of instruction, 11.

II. PREVIOUS STUDIES 13

Studies at the college level, 13. Studies at the secondary school level, 16. Relation of studies to book, 21.

PART II

STUDIES IN THE MEDICAL SCHOOL

III. TECHNIQUES IN THE TEACHING OF HUMAN ANATOMY 25

Statement of problem, 25. Plan of study, 26. Grouping of students, 27. Plan of instruction, 31. Measuring instructional outcomes, 32. Fall quarter data, 33. Winter quarter data, 36. Individual students both quarters, 41. Ratings of anatomy 1925-26 for comparison, 47. Prediction data, 49. Discussion of findings, 50. Faculty opinion, 52. Student opinion, 53. Ratings in subsequent courses, 53.

CHAPTER PAGE

IV. Laboratory versus Library in the Teaching of Human Physiology 55
Problem, 55. Grouping of students, 56. Achievement ratings, 59. Prediction data, 62.

V. Effect of Reduction of Time Given to Laboratory in Human Physiology 64
Problem, 64. Equation of groups, 65. Achievement ratings, 68. Subsequent data, 70.

PART III

STUDIES IN THE DEPARTMENT OF PHYSICS

VI. Preliminary Testing in Mechanics 75
Preliminary tests, 75. Value of scaling individual test items, 78. Scaling of preliminary tests, 79. Validity and reliability of preliminary tests, 83. The achievement of students in high school physics, 84.

VII. Measures of Achievement in Mechanics 89
Achievement tests, 89. Comparisons of Sets A and B, 93. Manner of scoring tests, 94. Validity, reliability, and significance of tests, 96. Student preference for tests, 99. Achievement of college groups, 100.

VIII. Pairing Students for Educational Experiments 102
Criteria for equation, 102. Use of correlations in pairing students, 105.

IX. Group Comparisons in Mechanics 116
The architectural group, 116. The architectural engineering group, 118. Other college groups, 120. Comparisons for 1923-27, 123.

X. Student Opinion in Mechanics 126
Value of student opinion, 126. Questionnaire to students in mechanics, 126. Opinion on the course in mechanics, 127. Values of laboratory, 128. Features of instruction, 129. Test preference, 130. Discussion of questionnaire returns, 132.

CHAPTER PAGE

XI. The Development of Prognostic and Achievement Tests in Mechanics 136
Technique, 136.

XII. An Experiment to Determine the Effect of Class Size in Physics 139
Problem, 139. Plan of experiment, 139. The final objective test as a criterion of achievement, 140. Comparisons of groups in large and small classes, 142. Relationships among ratings, 146.

PART IV
CONCLUSIONS

XIII. Recommendations and Conclusions 151
Purpose of studies, 151. Logic of studies, 152. Criteria of judgment, 153. A suggested technique, 154. Conclusions, 155.

PART V
A SUPPLEMENTARY STUDY IN THE TEACHING OF COLLEGE PHYSICS

XIV. Some Possible Determinants of Achievement in College Physics 161
Class size, 161. Bases of pairing students, 162. Comparisons of groups, 163. Testing, 166. Previous courses in the secondary school, 170. Knowledge of high school physics, 173. Achievement in college physics, 180. Interest in science, 181. Conclusions, 183.

TABLES

TABLE		PAGE
I.	Summary of Data for the Equation of Groups in Anatomy, Fall Quarter, 1926-27 (Pairs Only)	29
II.	Summary of Data for the Equation of Groups in Anatomy, Fall Quarter, 1926-27 (All Students)	30
III.	Distribution of Letter Marks in Anatomy, Fall Quarter, 1926-27 (All Students)	33
IV.	Distribution of Letter Marks in Anatomy, Fall Quarter, 1926-27 (Pairs Only)	33
V.	Summary of Achievement Ratings (Numerical) in Anatomy, Fall Quarter, 1926-27 (Pairs Only)	34
VI.	Summary of Achievement Ratings (Numerical) in Anatomy, Fall Quarter, 1926-27 (All Students)	35
VII.	Summary of Data for the Equation of Groups in Anatomy, Winter Quarter, 1926-27 (Pairs Only)	36
VIII.	Summary of Data for the Equation of Groups in Anatomy, Winter Quarter, 1926-27 (All Students)	37
IX.	Summary of Achievement of Groups in Anatomy, Winter Quarter, 1926-27 (Pairs Only)	38
X.	Summary of Achievement of Groups in Anatomy, Winter Quarter, 1926-27 (All Students)	39
XI.	Distribution of Letter Marks in Anatomy, Winter Quarter, 1926-27 (All Students)	40
XII.	Distribution of Letter Marks in Anatomy, Winter Quarter, 1926-27 (Pairs Only)	40
XIII.	Summary of Relative Ratings of Students in Anatomy, 1926-27 (In Four's, Winter Quarter)	42
XIV.	Summary of Relative Ratings of Students in Anatomy, 1926-27 (In Four's, Fall Quarter)	42
XV.	Summary of Relative Ratings of Students in Final Written Test (In Four's, Winter Quarter)	43
XVI.	Summary of Relative Ratings of Students in Final Written Test (In Four's, Fall Quarter)	43
XVII.	Summary of Relative Ratings of Students in Final Practical Test (In Four's, Winter Quarter)	43
XVIII.	Summary of Relative Ratings of Students in Final Practical Test (In Fours, Fall Quarter)	44
XIX.	Summary of Relative Ratings of Students in Quiz Averages (In Four's, Winter Quarter)	44
XX.	Summary of Relative Ratings of Students in Quiz Averages (In Four's, Fall Quarter)	44

TABLE		PAGE
XXI.	Summary of Relative Ratings of Students in Laboratory Work (In Four's, Winter Quarter)	44
XXII.	Summary of Relative Ratings of Students in Laboratory Work (In Four's, Fall Quarter)	45
XXIII.	Summaries of Tables XIII to XXII (Inclusive)	45
XXIV.	Distribution of Letter Marks in Anatomy, 1926-27, Counting Only Those Who Reversed	47
XXV.	Correlations between Final Ratings in Anatomy and Various Criteria	49
XXVI.	Correlations between Other Tests	50
XXVII.	Number of Times Sections A Exceed Sections B in the Mean Ratings in All Courses and Vice Versa	54
XXVIII.	Summary of Group Comparisons in Physiology	56
XXIX.	Reliabilities of Differences between Sections A and B in Physiology	57
XXX.	Summary of Data for the Equations of Groups, Sections A and B in Physiology (Pairs Only)	58
XXXI.	Reliabilities of Differences between Sections A and B in Physiology (Pairs Only)	58
XXXII.	Summary of Achievement Ratings of Students, Sections A and B in Physiology (All Students)	59
XXXIII.	Reliabilities of Differences between Sections A and B in Physiology (All Students)	60
XXXIV	(a). Comparisons of Achievement Ratings, Sections A and B in Physiology (Pairs Only)	61
	(b). Comparisons of Achievement Ratings, Sections A and B in Physiology (Twenty Pairs)	61
XXXV.	Reliabilities of Differences, Sections A and B in Physiology (Pairs Only)	62
XXXVI.	Coefficients of Correlation	62
XXXVII.	Summary of Data for the Equation of Groups, Sections C and D in Physiology (Pairs Only)	65
XXXVIII.	Reliabilities of Differences between Sections C and D in Physiology (Pairs Only)	65
XXXIX.	Comparisons to Show How Nearly Pairs in Section C Are Random Samplings of That Section	66
XL.	Comparisons to Show How Nearly Pairs in Section D Are Random Samplings of That Section	66
XLI.	Summary of Data for the Comparison of Groups, Sections C and D in Physiology	67
XLII	(a). Comparisons of Achievement Ratings, Sections C and D in Physiology (Pairs Only)	68
	(b). Comparisons of Achievement Ratings, Sections C and D in Physiology (Twelve Pairs)	68

TABLE		PAGE
XLII	(c). Reliabilities of the Differences between Sections C and D in Physiology (Pairs Only)	69
XLIII	(a). Distributions of Scores in Hurd Physics Tests (First Test Composite) (All Students)	76
	(b). Distributions of Scores in Hurd Physics Tests (Second Test Composite) (All Students)	76
	(c). Comparisons of Preliminary Test Scores in Mechanics of Students in Different College Departments (First and Second Tests Combined)	77
XLIV.	Summary of Percentile Scores in Freshman Mental Test..	78
XLV.	Computation of S. D. Percentiles, Test X, Electricity and Magnetism	80
XLVI.	First Test Given to Entering Classes in Physics	82
XLVII	(a). Distribution of Scores in Hurd Physics Tests and a Combination of Several Tests	85
	(b). Distribution of Scores in Hurd Physics Tests and a Combination of Several Tests	86
XLVIII.	Comparisons of Students Who Had High School Physics with Those Who Did Not	86
XLIX.	Summary of the Distribution of Percentile Scores in the Freshman Mental Test	87
L.	Summary of the Distribution of Combined Scores in Hurd Physics Tests	87
LI.	Summary of the Distribution of Percentile Scores in the Freshman Mental Test	87
LII.	Distribution of Marks (Numerical) in Mechanics and M Values	89
LIII.	The Buchta Test Scaled According to the McCall T Score Technique	90
LIV.	Second Objective Test Scaled with Distribution of Raw Scores and M Values	91
LV.	Distribution of the Final Objective Test, as Scored Originally in Mechanics, and the Scaled M Score Values	92
LVI.	Summary of the Distribution of the Buchta Test When Given in Sets A and B	93
LVII.	Summary of the Distribution of the Second Objective Quiz When Given in Sets A and B	93
LVIII.	Ranking of 100 Students in the Final Objective Test on Original and Revised Scores	94
LIX.	Correlations between Halves of Achievement Tests in Mechanics	96
LX.	Correlations Showing Validity of Achievement Tests in Mechanics	97

TABLE		PAGE
LXI.	Comparisons of Individual Ratings in Final Grades and Scores on Buchta Test	98
LXII.	Comparisons of Individual Ratings in Final Grades and Scores on Final Objective Test	98
LXIII.	Students' Preference for the New Type of Test versus the Conventional Type	99
LXIV.	Summary of Final Achievement Ratings in Mechanics, to Compare Those Who Have Had High School Physics with Those Who Have Not	99
LXV.	Summary of Engineering, Chemistry, and Academic Students in Achievement Ratings in Mechanics	100
LXVI.	Comparisons of Achievement Ratings in Mechanics of Various College Groups	100
LXVII.	Summary of Distributions of Freshman Percentile Intelligence Ratings	101
LXVIII.	Correlations Obtained between Final Achievement Ratings in Mechanics and Other Past Records, Respectively	103
LXIX.	Table to Show Scaled M Values of Letter Marks A, B, C, D, E, and F, in Various Courses for Students Enrolled in Mechanics	104
LXX.	Records of Architectural Students	117
LXXI.	Records of Paired Students in Architectural Engineering for Comparison with Architectural Students	118
LXXII	(a). Comparisons to Show the Significance of Differences in the Means of the Ratings of Architectural and Architectural Engineering Students	119
	(b). Number and Per Cent of Students in Architectural Engineering Exceeding Their Paired Companions in Architecture	120
LXXIII.	Summary of Records of Paired Students for Purposes of Comparison	121
LXXIV.	Percentages of Students in Each Group Exceeding Their Paired Companions in Architecture	122
LXXV	(a). Summary of the Distribution of Final Numerical Marks of Students in Mechanics, Years 1923-27, Inclusive	124
	(b). Reliabilities of Differences between Mean Ratings of Various Groups in Mechanics, 1923-27, and of Students in Architecture	124
LXXVI.	Distribution of Letter Ratings in Mechanics	132
LXXVII	(a). Summary of the Ratings of Students in Large Section in Heat	143
	(b). Summary of the Ratings of Students in Small Section in Heat	144

TABLE		PAGE
LXXVIII.	Summary of Ratings of All Students Enrolled in Heat and Those of the Large and Small Group Sections	145
LXXIX	(a). Differences in Means of Paired Groups, Large and Small Sections in Heat	146
	(b). Differences in Members of Paired Groups, Large and Small Sections in Heat (Nine Pairs)	147
LXXX	(a). Distribution of Marks in Mechanics Given to the Members of the Group Enrolled in Heat	148
	(b). Transmutation of Final Objective Score in Heat into Letter Marks	148
LXXXI.	Ratings Used for Determining Best Means of Pairing Students in Large and Small Classes in Magnetism and Electricity and Correlations with the Final Rating in M Scores	162
LXXXII	(a). Comparisons of Large and Small Classes in Magnetism and Electricity	163
	(b). Comparisons of the Means of the Small and Large Class Ratings in the Criteria of Table LXXXI	165
LXXXIII.	Comparisons of Achievement Ratings of Large and Small Groups Given in the Criteria of Tables LXXXII (a) and LXXXII (b)	165
LXXXIV.	Distribution of Final Conventional Test Scores and M Values in Magnetism and Electricity	168
LXXXV.	Distribution of Scores in Final Objective Examination in Magnetism and Electricity	169
LXXXVI.	Coefficients of Correlation with Final Ratings in Magnetism and Electricity Which Might Have Been Considered in Pairing Students for This Experiment	169
LXXXVII.	Distribution of Ratings in Magnetism and Electricity	170
LXXXVIII.	Distribution of Scores in Hurd Test X of Students Enrolled in Magnetism and Electricity	171
LXXXIX.	Distribution of Scores in Hurd Test X in Magnetism and Electricity (High School and Non-High School Physics Students)	172
XC.	Items in Preliminary Test X Answered Correctly by 90 Per Cent of Students in Magnetism and Electricity	173
XCI.	Items in Preliminary Test X Failed by at Least 50 Per Cent of Students in Magnetism and Electricity	174
XCII.	Items in Preliminary Test X Failed by at Least 75 Per Cent of Students in Magnetism and Electricity	176
XCIII.	Items in Preliminary Test X Answered Correctly by 90 Per Cent of Students in Magnetism and Electricity Having Had High School Physics	177

TABLE PAGE

XCIV. Comparisons of Percentages of Students Giving Correct Responses to Items in Preliminary Test X 178

XCV. Items in the Final Objective Examination in Magnetism and Electricity, with Number and Per Cent of Students Making Correct Responses to Each Item 180

XCVI. Means and Standard Deviations of Certain Ratings in Physics for Students Having Taken Various Numbers of High School Science Courses 182

XCVII. Means and Standard Deviations of Final Ratings in the University Course in Magnetism and Electricity for Students Taking Different Courses in University Physics 183

CORRELATION CHARTS

CHART		PAGE
I.	Numerical Grades (x) and Intelligence Percentiles (y) in Mechanics	106
II.	Numerical Grades (x) and Preliminary Test—Total Raw Score (y) in Mechanics	107
III.	Numerical Marks (M Scores) (y) and Preliminary First Test (M Scores) (x) in Mechanics	108
IV.	Numerical Grades (M Scores) (x) and First Preliminary Test (M Scores) (y) in Mechanics	109
V.	Numerical Grades (M Scores) (x) and Honor Points per Credit (y) in Mechanics	110
VI.	Numerical Grades (M Scores) (x) and Chemistry Letter Grades (M Scores) (y) in Mechanics	111
VII.	Numerical Grades (M Scores) (x) and Freshman Mathematics Letter Marks (M Scores) (y) in Mechanics	112
VIII.	Numerical Grades (M Scores) (x) and Freshman Mathematics Letter Marks (M Scores) (y) in Mechanics	113
IX.	Numerical Grades (M Scores) (x) and Freshman English Letter Marks (M Scores) (y) in Mechanics	114

PART I

TEACHING PROBLEMS

INTRODUCTION

The achievements of science are the greatest boast of our present age. Extravagances of our fathers have been transformed by science into commonplace necessities of life, and new and undreamt-of comforts and luxuries have been added. Everybody takes science for granted in medicine, business, engineering, agriculture—in the daily walks of life.

Nowhere has science come into a more dominant place than in our institutions of higher learning, where it was cradled if not actually born. Certain it is that these institutions are the greatest source of scientists—the training ground for increasing thousands who annually go out to solve with the techniques of the physical and biological sciences the theoretical and practical problems of society. No more fundamental contribution could be possible to science and to society than this same function of teaching and training scientists. Are we doing a good job of this teaching? It may be but it is not necessarily so.

The scientific method is common to all sciences. The investigator in human diseases, for instance, relies on accurately described observations and controlled experiments to learn the origin of a specific disease. The physicist, chemist, and biologist use the same method in their fields. Facts in science are not taken for granted—they must be proven, and that proof is evaluated in the light of the criticism and experiments of other scientists. On this fundamental question of determining and proving facts there is no disagreement among scientists.

Since the scientific method has been so universally accepted, it has also been quite generally taken for granted that all of the scientists' activities are equally sound in value—a most unscientific conclusion. Our science teachers in colleges and universities have a dual rôle—they are first, scientists; and second, teachers. Do they apply the same scientific method to their teaching that they use in their investigation in science? Have they an experimental attitude toward classroom procedures? There is not much evidence that this is the case. Certainly the published scientific experiments are few in number.

Perhaps the best illustration of this anomalous situation is found in the commonly accepted use and the prevailing estimate of the value of the laboratory method in the teaching of science. Now the laboratory method has been a most effective one for advancing our knowledge in the sciences. It is unquestionably the foundation stone of scientific investigation. But does it necessarily follow that success in the teaching of science can only, or even chiefly, be obtained by the laboratory method? That such a method of teaching science is actually superior requires scientific proof.

Modern conditions in our colleges and universities have brought rapidly increasing demands on the time and resources of teachers of science. The tremendous expansion of human knowledge has increased and complicated the curricular requirements of colleges. Large bodies of students and necessary economies in administration and teaching resources call for close scrutiny of expenditures and the greatest possible efficiency. Science teachers are becoming acutely aware of the fact that savings in time and resources gained through cheaper and more effective teaching may leave greater opportunities for research.

But methods of teaching are not the only concern of the science teacher who seeks scientific proof of his success as a teacher. A multitude of problems meet the enquiring mind. What are the aims and objectives of any given course of science? Chemistry courses for the chemical engineer may not be suitable at all for a student majoring in English. Can science courses be more exactly and profitably adjusted to the future occupations of students?

The curriculum in science offers a host of problems for careful study. Do prerequisite courses actually function as such? Are we guilty of unnecessary duplication or overlapping in the science curriculum?

Perhaps even more fundamental problems appear in the consideration of measures of achievement among students in our science courses. Does the science teacher obtain the results he thinks he gets or claims for his course? It may be comparatively simple to test by objective methods the accumulation of facts by the student, but can the science teacher be sure that he has actually taught the student a scientific attitude or good judgment? Have

such claims been measured? Can they be measured? It would seem to be the concern of an advocate of the scientific method to investigate and determine the facts in these matters. Certainly a teacher of science is a teacher as well as a scientist. Is he scientific in both functions or merely in one?

Costs of education are no longer merely the concern of the college administration. The rising tide of educational expense to teach and train the hordes of college students in the rapidly increasing and highly specialized professions demands the cooperation of teachers as well as the generosity of legislatures and philanthropic providers of educational funds. Are we getting our money's worth in our teaching? Can we effect economies without injury to our teaching? The meager investigational results so far obtained indicate clearly that such economies are at least possible.

Time was when the problems of teaching were lodged in the departments and colleges of education. On the one hand, educationists assumed the rôle of final critics in the art of teaching. Teaching was a profession only casually related to its subject matter. On the other hand, science teachers, secure in their knowledge of their own science, ignored the dogmas of education and glorified the subject matter of science. To them methods were inconsequential and educationists, meddling theorists.

But times have changed. The modern educationist has espoused the scientific method. He, too, is a scientist practicing the same rigid rules and procedures of other scientists. He has proven his sincerity by challenging and investigating his own methods and results. He has won the right to challenge other teachers, and he has not overlooked the scientist who perhaps, above all others, must defend himself and justify his results as a teacher by the methods of science. Here and there science teachers have accepted the challenge and acknowledged a responsibility.

The investigation of science teaching is a joint responsibility. The physiologist whose research problem juts out into the field of physics freely seeks the cooperation of physicists. Medical colleges employ bio-chemists and bio-physicists. Just so may the science teacher recognize the need of cooperation with educational investigators, and the educationist appreciates the need of cooperation from the subject-matter specialist. It follows clearly, there-

fore, that the arena for the investigation of the college-teaching problems in science lies in the field of that science and that the research in these problems can best be attacked by joint efforts of the science teacher and the educational cooperator schooled in the investigation of teaching problems.

Here, then, is a clear-cut opportunity for most important contributions, not merely to the science of education, but also to any subject-matter science that consents to investigate and understand more clearly its educational possibilities and responsibilities. The objectives of educationists who investigate science teaching and scientists who investigate teaching methods are identical. Destructive antagonisms must give place to productive cooperation.

In the light of such considerations the Committee on Educational Research of the University of Minnesota organized a few years ago a subcommittee on research in the teaching of science in the University. It is significant of the new attitude that this committee of nine members represented six colleges and nine departments of instruction in pure and applied, elementary and advanced, specialized sciences, with only one member from the College of Education. They had one objective in common: to investigate their own teaching methods and procedures by means of their own methods of investigation.

In a preliminary canvass of the possible fields of investigation, it became clear that a large array of problems were calling for solution. Since college science teachers are usually kept busy by routine teaching and research duties, voluntary contributions of time and opportunities by individuals and departments were obviously the only available avenue of progress. Teachers confronted by immediately pressing problems or teachers particularly interested in specific questions in science teaching presented the possibility of an immediate attack. Studies in the teaching of courses in human anatomy and physiology and in the field of physics were soon under way.

The committee on research in the teaching of science has interpreted its primary function as one of coordination, cooperation, and stimulation in this voluntary attack upon any phase of the problems presented. It has desired to encourage investigation wherever such might be properly undertaken, and to consolidate the

efforts of individual investigators into a comprehensive program which, with the passing years, might be expected to shape itself into a more or less orderly attack on the whole field.

One of the first desirable steps in the program seemed to be a preliminary analysis of the more obvious fields of investigation involved in the teaching of college science. A suggested comprehensive program was therefore prepared and presented to all science departments of the University in the hope that such a program would clarify the situation and encourage teachers of science here and there to attack specific problems of peculiar interest to them. This comprehensive program has been published in some detail in *Problems of College Education* (University of Minnesota Press, 1928). It was, of course, made clear by the committee that it had no intention of starting investigations in all, or even in many, of the possible fields of research. The program was merely suggestive.

The nature of the analysis can be told best in a brief summary of the more important divisions.

1. The curriculum in science presents problems of analysis of existing conditions: studies of prerequisite courses, their function and effect on sequent courses; and problems of correlation of existing courses in the interests of greater efficiency and usefulness.

2. Aims and objectives of instruction in science involve consideration of science courses in relation to social and industrial life. Among others this includes studies of the historical setting, of student personnel, of alumni occupations, of changing environments, and the growth of science.

3. Methods of instruction in science open up a vast array of problems of particular and immediate interest. Analysis of outstanding success of present and past teachers would serve to clarify our present values. Of great importance are the possible experimental investigations of existing or suggested methods of teaching college science. It is assumed that such investigations will be by scientific methods and may deal with any of the numerous standard methods or combination of methods of teaching, or they may deal with methods entirely new.

4. Measurements of achievement in science involve particularly intricate problems for investigation. A preliminary survey elicited

opinions from teachers of elementary science which credited forty-three kinds of values to science teaching. But even the accumulation of facts in a science course has not been submitted to any uniformly comparable method of evaluation. If the remaining numerous values are actually founded on facts, one may properly demand units of measurement and actual accomplishment in terms of such units. Research in this field may yield results of fundamental importance.

5. Costs of instruction in science are no longer merely the concern of administrators. Such cost is admittedly high, especially in laboratory and analogous methods. The justification of such costs can be demonstrated only by a greater efficiency in teaching. Accurate and quantitative studies are essential for a satisfactory scientific answer to these problems.

Several investigations at the University of Minnesota bearing on some phase of the teaching of science have already been published through the University of Minnesota Press *(Problems of College Education)*. Other investigations are under way. Research in general, and particularly that supported largely or entirely by voluntary effort, is necessarily slow and cannot profitably be hurried.

Among the most exhaustive and most comprehensive studies in the Minnesota program have been those carried on by the Departments of Anatomy, Physiology, and Physics in cooperation with Dr. A. W. Hurd, executive secretary of the committee on research in teaching science. The chief results to date are embodied in this volume by Dr. Hurd. The book is not, however, merely a record of specific contributions, valuable as these may be in their own peculiar field. It is a contribution to the methods of research in the general field of science teaching. In a sense it is a pioneer's trail and exhibits the difficulties and pitfalls inherent in this somewhat new field of scientific investigation. It is hoped that the reader will appreciate the significance of this path-breaking character of the contribution.

There is one additional phase of research in the teaching of science embodied in this volume which ought not to pass unnoticed. A sincere desire is evident on every hand among colleges and universities to improve the quality and effectiveness of classroom in-

struction. Various and numerous devices are in use to bring about such improvement. Lectures, seminars, and faculty discussion groups are merely a few of the many on trial. In my opinion none of these is fraught with more promise of success than the very method exemplified in this volume. Nothing appeals more strongly to the young and ambitious teacher of science than a challenge to him to prove by his own scientific method the results of his job as an educator. He is both trapped and intrigued. The least he can do is to study his own methods in the light of other admittedly good teaching to justify his own procedure. If he is truly scientific, he will sooner or later see to it that his teaching is founded on a sound scientific basis either of his own making or contributed by the efforts of others. If his life is truly devoted to science and the teaching of science, he cannot escape the issue. Rather will he confidently and resolutely meet it by conscientious effort to improve his teaching and to base that improvement as he bases his subject-matter science on the results of sound research.

E. M. Freeman, *Chairman*
Subcommittee on Teaching of Science

CHAPTER I

NEED FOR EXPERIMENTATION

Some work, even though crude and elementary in character, has been done in educational experimentation in the teaching of science at the secondary school level. Comparatively little, however, has been done at the college level. Indeed, the educational world has not yet emerged from the Aristotelian age of science in so far as the settlement of educational problems is concerned. The opinions of those in charge of educational work are still largely the bases for the solution of the most pressing problems. Authority is the guide, and the loudest talker or the most plausible debater often becomes the constituted authority. But the time is near, when the procedures of Galileo must become the bases for deciding educational questions. For several hundred years the logic of Aristotle determined whether heavy bodies fall faster or slower than light bodies. The experimental technique of Galileo served to prove the logic of Aristotle unsound. Only experimentation can satisfactorily settle many of our educational disputes.

The difficulties in educational experimentation are by no means small. Measurements seem intricate because the techniques are new and the field largely undeveloped. But Ohm utilized the work of his predecessors and made clear the relations existing between electric currents (amperes), pressures (volts), and resistances (ohms), even though the nature of electricity was unknown. In the field of education workers are now needed to analyze and isolate quantities which can be used better to measure educational factors and products and to make clear existing relationships. It is not essential that the exact nature of learning be known.

It will be of interest, first of all, to note the chaotic state of opinion in the minds of science teachers in regard to one of the most firmly fixed tenets of science teaching, namely, laboratory work. It is the most universally emphasized feature of instruction in science and never fails to receive major attention in any discussion. Yet the variety and confusion of opinion as to objectives in laboratory work may be illustrated by the first step taken in preparing this study.

In the early part of the school year 1926-27 a questionnaire was circulated among teachers of science at the University of Minnesota. Each person was asked to state specifically what he believed to be the particular function of the laboratory in instruction in science. Questionnaire Summary I is based on the replies received. It illustrates not merely the many and various opinions but also the difficulties in an attempt to verify experimentally all or any of them. The summary attempts to give them in essentially the same form as received.

Questionnaire Summary I

FUNCTION OF THE LABORATORY IN SCIENCE

Results of Questionnaire to Faculty Members in Science, University of Minnesota

(35 Usable Returns)

Statements and Number of Times Mentioned

1.	Actual practice in the performance of operations in the field of the subject	8
2.	To develop the technique of the subject	8
3.	To develop manipulative skill	7
4.	First-hand contact for acquaintance	7
5.	Makes ideas realistic or concrete	5
6.	Develops interest	4
7.	To aid memory	4
8.	Application of principles	4
9.	For demonstration	3
10.	To give the scientific manner of thought and training in drawing conclusions	3
11.	Gives opportunity for developing sense perception and acquisition of concepts	3
12.	Develops powers of observation	3
13.	For illustration	2
14.	Develops ability to do independent work	2
15.	Instructs by making visual impressions	2
16.	Gives a chance for student initiative	2
17.	Gives accurate concepts	1
18.	Clarifies the textbook	1
19.	Supplements the text in clarifying principles	1
20.	Repetition by concrete contact	1
21.	Teaches correct proportions and proper arrangement of the parts	1
22.	Furnishes experience on which to base reasoning	1

23. Coordinates theories and facts by interpretation of observations..... 1
24. Teaches a student self-confidence..... 1
25. Develops an inquiring mind..... 1
26. Develops the quantitative and experimental viewpoint..... 1
27. Develops ability to obtain and evaluate first-hand information..... 1
28. Develops ability to discard the immaterial..... 1
29. To show a method of ascertaining facts and answering questions..... 1
30. To create confidence in the text..... 1
31. To provoke thought..... 1
32. To suggest problems..... 1
33. To give opportunity for developing inventive ability..... 1
34. To show the difficulty of generalization..... 1
35. First-hand acquaintance with the experimental side..... 1
36. Overcomes disgust (animal dissection)..... 1
37. Develops self-control 1
38. Trains a student in planning, in care, accuracy, and the use of data and weighing of evidence..... 1
39. Shows a student whether he is really interested in, or properly prepared for, scientific laboratory work..... 1
40. Gives opportunity for analysis and comparison..... 1
41. Develops synthetic ability..... 1
42. To get a thorough and systematic knowledge..... 1
43. Makes the student study the text in order to write up the experiments 1

The statements indicate various points of view. One instructor is endeavoring to make experts of his students in his own particular specialized subject. One is trying to develop the inherent powers of his students in observation, scientific manner of thought, and initiative. One is attempting to give his students general information in the specialized-subject field. These points of view imply aims and objectives that may conflict in any given course and should, consequently, be more carefully differentiated. In determining the function of laboratory work, therefore, it is evident that specific studies in specific situations are desirable. The function depends upon the situation.

The second questionnaire results are those received from the teachers of science in secondary schools who were enrolled in certain educational courses in natural science at the University of Chicago, Summer Session, 1927. They are the opinions of these students on the values they, themselves, believed that they had received from certain features of courses pursued by them in science. The

most valuable feature in their minds, judged by the weighted-average choice,[1] is the lecture-demonstration, while the least valuable is the recitation. It may be noted that, judged by first choices alone, the laboratory is most popular, but, by the method of weighted average, the lecture-demonstration ranks just a little higher.

Questionnaire Summary II

VALUES OF CERTAIN FEATURES OF INSTRUCTION IN SCIENCE

Results of Questionnaire to Teachers of Science Enrolled in Classes in the Teaching of Science, University of Chicago, Summer Session, 1927

(37 Usable Returns)

A. Values of lecture, lecture-demonstration, laboratory, textbook study, and recitation expressed by rank choice of 1 to 5, inclusive, judged from their experience as students in science.

Methods	Choices					Total	Weighted-Average Choice
	1	2	3	4	5		
Lecture	6	6	7	9	8	36	3.2
Lecture-demonstration	10	11	8	3	3	35	2.4
Laboratory	12	6	7	6	5	36	2.6
Textbook study	5	10	10	7	5	37	2.9
Recitation	4	4	5	11	13	37	3.7

B. Abilities Developed by Laboratory Work.

	Frequency of Statement
1. Manipulation	20
2. Thinking through problematic situations	12
3. Independent work	6
4. Following instructions carefully	6
5. Logical methods of recording data	6
6. Facility or accuracy in measurement	6
7. Ability to understand fundamental principles and facts	5
8. Gather materials together for carrying through problems	4
9. Observation of detail	4
10. Thinking in quantitative terms	3
11. Ability to keep a fairly good notebook	3

[1] The weighted-average choice was determined by multiplying each choice by its frequency and dividing the sum of these products by the number of persons. By using this method of computation each choice receives some consideration.

12. Ability to cooperate with others 3
13. Accuracy in getting data 2
14. Ability to waste time 2
15. Ability to work carefully 2
16. Scientific method 2
17. Verify textbook material 2
18. Ability to systematize 2
19. Enthusiasm 2
20. Habits of accuracy, patience, etc. 2
21. Skill in technique 2
22. Gain confidence 1
23. Devise new ways 1
24. Appreciation of others' work 1
25. Helped me to choose my vocation 1
26. Knowledge of electrical devices 1
27. Emotional serenity 1
28. Accuracy in drawing conclusions 1
29. Ability to express ideas 1
30. Honesty 1
31. Ability to picture theory actually working 1
32. Ability to see relations 1
33. Critical attitude 1
34. Persistence 1
35. Interpretation of printed instructions 1

C. Could these have been developed by some procedure other than laboratory?

Yes—12, No—19, Possibly—2.

D. Was time wasted?

Yes—31, No—4, At times—1.

These results from the two questionnaires are self-explanatory. They represent only opinion. They serve chiefly to illustrate the great variety of opinion on objectives and achievements in one of the most commonly accepted features of science teaching. The need for factual data is evident. Experimental education must be depended upon to furnish these data.

The following chapters contain accounts of experimental work at the University of Minnesota which attempts to attack educational problems from the viewpoint of the experimenter. Attention is directed particularly to the techniques of experimentation. Under the conditions existing they are worth serious consideration. It is hoped that they may at least be suggestive of further work of a similar character.

CHAPTER II

PREVIOUS STUDIES

Studies at the College Level

Colton[1] has given the first account, known to the writer, of an experimental procedure in testing methods of the teaching of science at the college level by comparing contrasted groups of students. In 1921-22 ninety-six premedical students taking general zoology at the University of Pennsylvania were grouped at random into four sections of twenty-four each. Each of two instructors taught two sections, one section by each method. There were thus two sections that were taking intensive work in the laboratory for training and two sections taking extensive work in the laboratory for information. The former covered little content, the latter, considerable. Different laboratory manuals were used in the contrasted sections. Many lectures and demonstrations were used to supplement the informational laboratory work, while very few were used with the "training" group. Both sections used the same text.

A final examination was used as the criterion of achievement. It consisted of the following items: (1) a diagram to show details in observation; (2) the same diagram scored for proportions and character of work; (3) a diagram of an imaginary section of an animal not studied, to measure constructive imagination; (4) a test of judgment consisting of (*a*) a statement containing errors, which were to be crossed out, (*b*) a paragraph containing facts and theories to be distinguished by double and single underlining, respectively, (*c*) a paragraph to be summed up in one sentence; (5) a problem in genetics to measure ability to reason from facts; (6) ten informational questions; (7) ten words to be used correctly; (8) the outline of a speech, to measure ability to organize; and (9) an estimate of morale by the instructor. There were no significant differences shown by the testing, though there were some differences on certain topics between instructors. Colton concludes that instructors should be allowed to use the laboratory method they like best.

[1] H. S. Colton, "Information versus Training; an Experiment in Laboratory Methods," *School Science and Mathematics*, 25:256-258.

A second study is reported by Bane[2] in courses in education and psychology. Five experiments were conducted to contrast results between a class-discussion method and a lecture method, respectively. One instructor taught all classes, using 8 to 16 teaching periods for each experiment. All experiments occupied a total of 158 periods of 50 minutes each and included 510 juniors and seniors in college.

In each experiment all factors that were thought to be effective in determining achievement were held constant with the contrasted sections. These were: (1) amount of teaching time, (2) subject materials (facts and general principles), (3) assignments (texts, references, problems), (4) objective tests (most of them true-false of known reliability), (5) intelligence (as measured by Forms A and B of Morgan's and Terman's tests of mental ability, respectively), (6) efforts of the instructor. Three methods were used in securing comparable achievement ratings: (1) totaling achievement quotients for each group, (2) adding scores of paired members of a section, and (3) adding scores of all members of a section whether paired or not paired.

Bane's conclusions were: (1) The lecture and class-discussion methods of college teaching appear about equally effective in the immediate recall of content material. (Three experiments favored the lecture and two, the class-discussion.) (2) The class-discussion is more effective than the lecture in delayed recall. (3) The lecture method is more suitable for the immediate recall of subject matter than for its retention to a later period, while the reverse is true of the class-discussion method.

Scheidemann[3] made a study of the "lecture-conference" method, as compared with an "individualized" method at the University of Iowa in 1925-26 in the field of educational psychology. Students enrolled for a six-semester-hour course, extending over one academic year, were divided into two groups by pairing them on the bases of (1) grade points made during the first collegiate year; (2) percentile ratings in the freshman entrance examination; and (3) grade

[2] Charles L. Bane, "The Lecture vs. the Class-Discussion Method of College Teaching," *School and Society,* 21:300-302.

[3] Norma V. Scheidemann, "A Comparison of Two Methods of College Instruction," *School and Society,* 25:672-674.

points and percentile ratings combined after conversion into comparable scores.

An attempt was made to keep instructional factors uniform except the method used, although it may be noticed that a different instructor taught each group. The instructors, however, were not aware that comparisons were being made. The same equipment was used for both classes; the aims of the two courses were the same; the content covered was the same; the same text was used; and the objective tests, devised by a distinterested person, were administered under uniform conditions to both groups. The experimenter concludes that the two methods, as followed, were of equal effectiveness.

Noll[4] in a Master's thesis (University of Minnesota, 1926), reports a study in college chemistry, comparing achievements of various college groups with varying times spent in the laboratory. Some of these groups had had high school chemistry and the others had not. Several bases of pairing students were used in succession, and achievements were measured by a specially constructed objective test and the Iowa Placement Examination, Series CT 1, Revised, A, Chemistry-Training. Noll drew only two final conclusions: (1) There are significant differences in achievement in the various groups. (2) "When two groups, which were almost identical in every respect, except in the amount of laboratory work, were compared as carefully as was found to be possible, the group having the greater amount of laboratory work showed some superiority in general achievement." This study is being continued.

Duddy[5] reports a study in a business course at the University of Chicago in 1923; Good,[6] one in the social sciences; and Persinger,[7] one also in the social sciences. None of these has received

[4] Victor Noll, *The Effect of Varying Amounts of Laboratory Work upon Achievement in General Inorganic Chemistry.*

[5] Edward A. Duddy, "Report on an Experiment in Teaching Methods," *Journal of Political Economy,* 32:582-603 (Oct. 1924).

[6] Carter V. Good, "An Experimental Study on the Merits of Extensive and Intensive Reading in the Social Sciences," *School and Society,* 23:758-760 (Dec. 12, 1925).

[7] C. E. Persinger, "Experiment in Quantity-Quality Assignment in College Classes," *Historical Outlook,* 17:293-295 (Oct. 1926).

the necessary statistical treatment. Crawford[8] reports a study in the results of college note-taking, following in general the rotation plan, as described by McCall.[9] Here, a true-false test was used for measuring purposes. Edmonson and Mulder[10] report a study on class size where army alpha scores were used as a basis for pairing students in classes in Education. Von Borgersrode[11] reports a very extensive study on class size, carried on in more than one hundred college classes at the University of Minnesota. Pairings of students were made on the bases of past school marks and mental, reading, and placement tests. For achievement ratings conventional marks and new type tests were used.

Studies at the Secondary School Level

The experimental studies described in the succeeding chapters have their models in studies made, especially, in the field of secondary education.

A study by Hunter[12] is one of the earliest mentioned in secondary school science, and has been summarized by Hurd[13] in his Master's thesis, 1924. Hunter seems to have been inspired by the work of Mayman.[14] Other studies mentioned by Hurd are those by Garber,[15]

[8] C. C. Crawford, "Some Experimental Studies of the Results of College Note-taking," *Journal of Educational Research,* 12:379-386 (1925).

[9] W. A. McCall, *How to Experiment in Education,* pp. 19-20.

[10] J. B. Edmonson and F. J. Mulder, "Size of Class as a Factor in University Instruction," *Journal of Educational Research,* 9: 1-12 (1924).

[11] Published as a Doctor's thesis and embodied in Hudelson's *Class Size at the College Level.* (University of Minnesota, 1928.)

[12] George W. Hunter, "An Experiment in the Use of Three Different Methods in the Class Room," *School Science and Mathematics,* 21: 875-890 (1921).

[13] A. W. Hurd, A Topical Method Compared with a Problem Method in the Subject of "Heat" in Physics (Master's thesis, University of Minnesota, 1924).

[14] J. E. Mayman, *Teaching Elementary Science in Elementary Schools,* New York Department of Education, Division of Research, Publication 13 (1915).

[15] Elinor Garber, "The Project Method in Teaching Chemistry," *School Science and Mathematics,* 22: 65-73 (1922).

Kiebler and Woody,[16] Bing,[17] Holroyd,[18] and Cooprider.[19]

Downing[20] has summarized experiments by Wiley, Cunningham, Phillips, Cooprider, Kiebler and Woody, and Anibel, which are of particular interest because of their findings in regard to the value of laboratory work. The general conclusion drawn from these is that the lecture-demonstration method "yields better results than the laboratory method in imparting essential knowledge, and is more economical of time and expense." This conclusion is disputed by Riedel,[21] who claims that the data do not justify such conclusions.

Hurd[22] gives an account of an experiment in high school physics, with comparisons of a topical and a problem method as used with two equated groups of pupils. Groups were originally equated on the basis of scores in the Terman Group Mental Ability Test, Form B, but the groups were compared, as well, in past teachers' marks, chronological age, mental age, I.Q., ratings in industry given by former teachers, and preliminary tests in the content of the course to be pursued. Achievement was estimated from repetitions of the preliminary test, both total final scores and gains being used as criteria. The more exhaustive procedures in equating and studying groups, and in determining achievement, together with reliability data, are outstanding features of this study.

[16] E. W. Kiebler and Clifford Woody, "The Individual versus the Demonstration Method of Teaching Physics," *Journal of Educational Research,* 7:50-58 (1923).

[17] Ada M. Bing, Measurement of the Results of Instruction of High-School Freshmen in English Composition (Master's thesis, University of Minnesota, 1922).

[18] Ina E. Holroyd, "An Analysis of an Experiment in Teaching First Year Mathematics," *School Science and Mathematics,* 22:114-121 (1922).

[19] J. L. Cooprider, "Oral versus Written Instructions and Demonstration versus Individual Work in High School," *School Science and Mathematics,* 22:838-844 (1922); and "Laboratory Methods in High School Science," *School Science and Mathematics,* 23:526-530 (1923).

[20] Elliot R. Downing, "A Comparison of the Lecture Demonstration and the Laboratory Methods of Instruction in Science," *The School Review,* 33:688-697 (1925); and *Teaching Science in the Schools* (1925).

[21] F. A. Riedel, "What, if Anything, Has Really Been Proved as to the Relative Effectiveness of Demonstration and Laboratory Methods in Science?" *School Science and Mathematics,* 27:512-519, 620-631 (1927).

[22] A. W. Hurd, *loc. cit.*

Curtis[23] has made a valuable summary of several experimental studies in the teaching of science, to which book the reader is directed. Carpenter[24] reports a comprehensive study in high school chemistry. The methods compared were: (1) demonstration method, (2) individual method, (3) the group-of-two method, and (4) the control method (the student takes the examination before he performs the experiment). This study was carried on in twenty-three schools in fourteen states with thirty-four classes involved. The rotation[25] plan was used, and the methods were compared by the ratings of the various groups on a specially constructed objective examination in chemistry. These ratings were expressed in sigma units.

The conclusions briefly summarized are: (1) The demonstration is equal to the laboratory method, judged by the measures used. (2) One student working on an experiment is better than two, though the difference is not great. (3) The intelligent student succeeds no matter what the method. (4) Students lower in intelligence seem to do better with the demonstration method. (5) Achievement in general is higher with higher intelligence.

Hurd,[26] in a second study in high school physics, compared a stereotyped teacher-assigned lesson plan and a novel pupil-assigned lesson plan. Equated groups were formed on the basis of: (1) ratings in sigma units in two mental tests; (2) chronological age; (3) I.Q.; (4) ratings in the first semester of physics (computed from the added scores of seven objective tests); (5) past marks in school work; and (6) ratings in industry given by former teachers. Preliminary tests in the subject content to be covered in the experiment were also used for comparisons. Achievement was measured by six objective tests, designed by the author, and by the Iowa Physics Test. Reliability data are included in the study.

The conclusions are: (1) The teacher-assigned lesson group

[23] F. D. Curtis, *Investigations in the Teaching of Science* (1926).

[24] W. W. Carpenter, *Certain Phases of the Administration of High School Chemistry*. Teachers' College, Columbia University Contributions to Education, No. 191 (1925).

[25] W. A. McCall, *loc. cit.*

[26] A. W. Hurd, "Subject Matter versus Character Traits," *Journal of Educational Method*, 5: 147-154 (Dec. 1925).

shows a significant superiority in solving problems. (2) The teacher-assigned lesson group is *probably* superior in the ordinary content of the course on Heat. (3) The pupil-assigned lesson group is practically equal to the teacher-assigned lesson group in information alone. (4) The pupil-assigned lesson group is practically equal to the teacher-assigned lesson group in the Iowa Physics Test in Heat.

Peters[27] compared a stereotyped textbook method with a reorganized method beginning with applications. Four high schools cooperated. McCall's[28] rotation method with one group and two experimental factors was used. Preliminary and final tests were given: one, testing fundamentals; the other, application. The data, as given, furnish evidence in favor of the reorganized method as used.

Knox[29] reports a study in high school chemistry on the relative merits of laboratory and demonstration methods. Four classes were used and two instructors, each instructor having one demonstration group and one laboratory group. All four classes were equated on the basis of mental ability. There were forty-two pupils in the demonstration group and forty-three in the laboratory group. Average sigma scores from three tests measuring immediate retention, delayed retention, and method of attack, give evidence favoring the demonstration group in all respects.

Johnson[30] compared lecture-demonstration, individual laboratory experimentation, and group laboratory experimentation. Pairings were made on the basis of mental-test scores for one series of experiments and on the added basis of first-semester achievement in the subject for the second series. Each group rotated, trying a

[27] C. J. Peters, "An Evaluation of a Reorganization of the Present Core of Subject Matter in High School Physics," *School Science and Mathematics*, 27:172-182 (1927).

[28] W. A. McCall, *loc. cit.*

[29] W. W. Knox, "The Demonstration Method versus the Laboratory Method of Teaching High School Chemistry," *School Review*, 35:376-386 (1927).

[30] P. O. Johnson, A Comparison of the Lecture-Demonstration, Individual Laboratory Experimentation, and Group Laboratory Experimentation Methods of Teaching High School Biology (Master's thesis, University of Minnesota, 1926).

third of the experiments by each method. Achievement was measured by objective tests. Some differences were obtained, but they are not significant. The numbers in the groups were small: eleven in each group in series one; and seventeen, sixteen, and seventeen, respectively, in the second series.

Duel[31] carried on an experiment in physics, comparing the "rotation" system of laboratory instruction, where one, or very few students, are working on different experiments at one time with the "even-front" system, where all are working on the same experiment at the same time. Pairings of pupils were made on the basis of mental-test scores, and objective tests were used to measure achievement. The "rotation" system showed significantly higher achievement and, in addition, was more economical in time and money. This experiment needs repetition with different instructors under different laboratory conditions, however, before final conclusions can be safely drawn.

Nash and Phillips[32] made a study of the relative value of three methods of teaching high school chemistry, viz., a "pupil method," an "instructor method," and a "combination method." Sixty-two pupils were distributed among three groups, the scores from the administration of the Miller Mental Ability Test, Form A, being used to form equated groups. Only fifteen pupils in each of the three groups were comparable, the remaining pupils being distributed at random. The only designation of the methods indicates that in the "pupil group" the individuals in the group worked by themselves; in the "instructor group" the instructor did the work; and in the "combination group" the two methods were combined.

Initial and final tests devised by the experimenters were used as measures of achievement. Both final scores and gains favor the instructor method, though the differences are small. The general conclusions suggest that methods must be adapted to the pupils and be dependent on the objectives of the course. This suggestion is based on a study of the individuals in each group.

[31] H. W. Duel, The Even Front versus the Rotation System in the Laboratory (Master's thesis, University of Minnesota, 1927).

[32] H. B. Nash and M. J. W. Phillips, "A Study of the Relative Value of Three Methods of Teaching High School Chemistry," *Journal of Educational Research*, 15:371-379 (May, 1927).

Hurd[33] has summarized several of his own investigations in a portion of a course in high school physics. The methods used at various times over the same field are an individual-project method (1925-26), a textbook-recitation method (1924-25), an individual-project method (1923-24), a textbook-recitation method (1923-24), a topical method (1922-23), and a problem method (1922-23). The same tests were used to measure achievement with all methods. The same instructor taught all classes reported. The classes were compared by average I.Q. as well as by achievement. The textbook-recitation method showed superiority, if anything, though differences are insignificant. The topical and problem methods, as used, showed somewhat lower achievement. The data suggest that different methods may actually produce different results. Most of our present tests are designed to measure the outcomes of a textbook-recitation method. The individual-project method, the topical method, and the problem method probably produce outcomes, which the usual tests do not measure. Therefore, when these methods are compared with the textbook-recitation method on the basis of ordinary test results, they do not show even equal achievements.

The experiments described in the succeeding chapters have their origins in the studies cited in this chapter. The citations have been made primarily to orient the reader in this type of study. The field of experimental education is still so new that there are many opportunities to investigate detailed phases. For example, much of the description in the succeeding chapters has been concerned with such specific phases as pairing of students, measuring achievement, validation of tests, finding the reliabilities of tests, studying the significance of tests, varying the techniques of planning experiments, and varying the techniques of comparisons. Attempts have been made to point out deficiencies in the techniques used, to suggest better techniques, and to make trials of as many techniques as may be feasible. Conclusions have been based only upon the evidence in hand. The reader should keep this viewpoint in mind.

[33] A. W. Hurd, "Suggestions on the Evaluation of Teaching Procedures in High School Physics," *School Science and Mathematics*, 27: 520-526 (1927).

PART II

STUDIES IN THE MEDICAL SCHOOL

CHAPTER III

TECHNIQUES IN THE TEACHING OF HUMAN ANATOMY

This experiment to determine the relative efficiencies of different groupings of students to a cadaver is one which was tried out during the fall quarter, 1926-27, in the Medical School and was repeated during the winter quarter of the same year. During the fall quarter certain groups of students in the anatomy laboratories were working in two's on a cadaver, while other groups were working in four's. The repetition during the winter quarter was for the purpose of shifting these groups. Those who worked in two's during the fall quarter were working in four's during the winter quarter and vice versa.

The Anatomy Department, for several years past, has followed the general practice of supplying one cadaver for every two students, each student working on the right or left half of the cadaver. Each student worked more or less independently, reading and following the directions from his own laboratory manual. With four to a cadaver (two on each side), on the contrary, one student read the directions while the other dissected the cadaver, each taking turns at reading and dissecting. This latter plan was a partnership affair with a division of the labor, each student doing his share but being allowed only one-half the time for dissection. With the former system, one student on one side of the cadaver did all of the dissecting on his half, having the opportunity at the same time of observing and talking with the student working on the other half. With four to a cadaver, two worked cooperatively on one half, while two others worked cooperatively on the other half. In this plan each of the four students had the opportunity of observing and talking with the others working on the same cadaver. Instructors' and students' opinions of the relative advantages and disadvantages of the two plans are given in the discussion toward the end of this chapter.

There are probably no generally accepted views as to the best practice to be followed in this matter. Undoubtedly, some departments of anatomy furnish one cadaver for groups of more than four

students. Especially is it necessary to do this when cadavers are not plentiful. Occasionally, the department at Minnesota has had to do with less than one cadaver to two students. The question of the effect of this on the efficiency of instruction arose in the mind of Dr. C. M. Jackson, the director, and this experimental testing of the results of instruction under the two sets of conditions was instituted. In other words, the director was not satisfied that students gained less knowledge of anatomy when working in groups of more than two to a cadaver, although his own choice was for two. He felt that there might be measurements of instructional results made which would indicate the great superiority of a smaller number of students to a cadaver, if there were such superiority. He wished to rest his final opinion, at least partially, on actual factual data. The purpose of the experiment was, then, to get as much factual data as feasible upon which to rest conclusions. The data are restricted to instructional outcomes, no material on administrative convenience or costs being provided. It is evident, however, that the saving in cadavers is approximately one-half with the four-to-a-cadaver technique.

It would seem to be a simple matter to get data on instructional outcomes; but a study of the following tables shows how really complicated it becomes. The statistical methods employed are not always easy for the uninitiated to understand, but are deemed essential to a study of this kind. The treatment in this study, in general, is as follows: (1) the consideration of the achievements of paired groups of students—one composed of students working in two's on a cadaver and one composed of students working in four's on a cadaver; (2) the consideration of the achievements of all students in the two groups even though not paired with other students; (3) the consideration of the achievement of each student working under one plan during one quarter compared with his own achievement working under the contrasted plan the succeeding quarter; (4) the consideration of achievements of individuals in a previous year in successive quarters, when the plan was the same for the two successive quarters; (5) the consideration of the achievements in succeeding courses in the Medical School of the students directly involved in this experiment. Attention is directed to the techniques of pairing students for the experiment and of

measuring and comparing achievements. There are no settled techniques yet established for work of this nature and less attention than is here given to details is not thought to be valid for drawing conclusions. There are undoubtedly opportunities for the establishment of more refined methods. A careful study of the tables will make clear the methods used. Only summaries are included. The reader is directed to the original and complete tables on file in the University Library for any desired additional information.

Grouping of Students for the Experiment

For the experiment as many pairs of students as it was possible to get were selected for two equated groups of students, equated on the bases of scores made in psychological tests and past ratings in school subjects. The scores made on the Miller test for graduate students,[1] and the average number of honor points made in required premedical subjects, exclusive of languages, were used as preliminary bases for pairing the students. Included in the original tables were other ratings which were given for the purpose of noting further the equality of the various pairs. It has been demonstrated that mental ability ratings and past school records furnish a fairly good prognosis of success in future school work, but whether the particular test or the particular past school records here used were sufficient for this purpose was not known. Therefore, ratings in Reading Test Sigma X[2] were included as well as average honor points per credit in all premedical subjects and marks in animal biology. The last named were made numerical by calling A (the highest rating), 1; B, 2; C, 3; D, 4; and F, 5. The past total premedical credits and required premedical credits were also included. Some dissimilarity of the pairs in these other criteria were noted. There are indicated, therefore, possibilities for improvement of methods of equating or pairing students for educational experiments of this kind.

The ninety-six students were finally divided into four sections, A, B, C, and D, each of which consisted of two divisions of twelve each. In Divisions 1, were put the members of the pairs who were

[1] W. S. Miller, University of Minnesota.

[2] University of Minnesota Reading Test.

to work in two's on a body. In Divisions 2, were put those who were to work in four's on a body. Each division contained a certain number of the members of the pairs. Thus there were eight pairs in Section A, Divisions 1 and 2; seven pairs in Section B, Divisions 1 and 2; five pairs in Section C, Divisions 1 and 2; and seven pairs in Section D, Divisions 1 and 2. There were twenty-seven pairs in all. The divisions, as first constituted, were filled in with other students chosen at random to make twelve in each division. By the end of the quarter two students had been eliminated, leaving ninety-four students in all. The twenty-seven pairs were, however, still intact.

All students were included in the records with the idea that perhaps there would be sufficient equality among the divisions, with all students counted, to make it possible to consider the total number as well as the pairs, i.e., that there might be equated groups working under the two plans, respectively, even though all members of one group were not actually paired with those of the other, individually.

Each section was placed under one instructor, with a different instructor for each section. Divisions 1 and 2 of each section were placed together in one of four laboratories. There were, therefore, twenty-four students in each laboratory, half of whom were grouped in four's and half in two's. One of the two students who dropped out during the quarter was from Division 1, Section A; the other from Division 2, Section C. Thus, in the tables, there are shown but eleven in each of the two divisions cited.

It was thought wise to consider the students (1) as pairs in divisions under a given instructor; (2) as pairs in divisions under all four instructors; (3) as total numbers in divisions under one instructor; and (4) as total numbers in divisions under four instructors. The pairs as individuals are also considered. Some of the original records are omitted in the summaries given in Tables I and II. In Table I, showing the summary of data used for selecting paired students, the ratings on industry are the means of two estimates given each student by two former instructors. These estimates were rendered numerical by giving A the highest rating, a value of 1; B, 2; C, 3; and D, 4; as with the marks in animal biology.

TABLE I

SUMMARY OF DATA FOR THE EQUATION OF GROUPS IN ANATOMY, FALL QUARTER, 1926-27*

(Pairs Only)

	No. Pairs	Miller Test	Sigma X Test	Av. H.P. Required	Animal Biology	Industry
Mean A_1—2b.....	8	70.62	65.00	1.53	2.42	1.94
S.D. A_1—2b.....	8	7.4	19.7	.31	.66	.39
Mean A_2—4b.....	8	71.37	67.62	1.54	2.12	2.19
S.D. A_2—4b.....	8	8.4	21.2	.32	.62	.50
Mean B_1—2b.....	7	57.57	62.71	1.43	2.74	2.14
S.D. B_1—2b.....	7	15.9	5.8	.34	.35	.35
Mean B_2—4b.....	7	59.43	60.43	1.42	2.55	2.43
S.D. B_2—4b.....	7	14.07	19.6	.34	.54	.62
Mean C_1—2b.....	5	65.80	68.00	1.67	2.33	2.00
S.D. C_1—2b.....	5	6.7	16.1	.19	.30	.84
Mean C_2—4b.....	5	69.00	69.80	1.66	2.00	2.00
S.D. C_2—4b.....	5	6.3	15.3	.17	.56	.55
Mean D_1—2b.....	7	70.71	75.00	1.78	2.02	1.79
S.D. D_1—2b.....	7	9.3	8.2	.47	.57	.25
Mean D_2—4b.....	7	70.86	79.86	1.79	2.08	1.86
S.D. D_2—4b.....	7	11.3	19.1	.47	.65	.69
Mean Total A—2b	27	67.48	67.37	1.59	2.38	1.96
S.D. Total A—2b	27	9.00	14.3	.38	.50	.49
Mean Total B—4b	27	68.67	69.70	1.59	2.16	2.13
S.D. Total B—4b	27	9.3	20.00	.38	.59	.63

* In this and succeeding tables letters A, B, C, and D indicate sections; $_1$, Division 1; $_2$, Division 2; 2b, two students to a body; 4b, four students to a body.

Statistical techniques described by Thorndike[3] were used to determine whether or not the differences between respective means and standard deviations for the two groups of students were great enough to be considered significant for the experiment, i.e., great enough to constitute grounds for belief that the groups should not be considered as equated groups. For the studies in the Medical School, the difference, divided by the probable error of the differences which is equal to 3, was used as a criterion of the significance

[3] E. L. Thorndike, *Theory of Mental and Social Measurements*, p. 190.

of a difference, i.e., if the ratio was less than 3 the difference was not great enough to be considered as a really true difference. It might be a difference due to the particular sampling of students in the experiment. If, however, the ratio was greater than 3, the difference was great enough to hold true for any similarly equated groups. Differences in means and standard deviations of the two groups for all ratings were computed and statistical significances determined. For the mean ratings of the groups of paired students, there were no significant differences. For the standard deviations there was but one significant difference, viz., that between the divisions of Section B in the Reading Test Sigma X. Even in this test there was no significant difference when the twenty-seven pairs in

TABLE II

SUMMARY OF DATA FOR THE EQUATION OF GROUPS IN ANATOMY, FALL QUARTER, 1926-27

(All Students)

	No. Students	Miller Test	Sigma X Test	Av. H.P. Required	Animal Biology
Mean A_1—2b	11	68.09	68.73	1.49	2.46
S.D. A_1—2b	11	10.17	18.71	.28	.54
Mean A_2—4b	12	68.92	70.67	1.67	2.04
S.D. A_2—4b	12	13.7	22.9	.43	.61
Mean B_1—2b	12	68.67	68.08	1.56	2.56
S.D. B_1—2b	12	13.7	13.1	.52	.52
Mean B_2—4b	12	57.42	56.50	1.41	2.64
S.D. B_2—4b	12	13.3	20.2	.28	.52
Mean C_1—2b	12	68.92	75.75	1.61	2.26
S.D. C_1—2b	12	10.6	15.8	.36	.57
Mean C_2—4b	11	68.00	77.18	1.56	2.15
S.D. C_2—4b	11	14.6	17.6	.31	.70
Mean D_1—2b	12	63.92	69.58	1.68	2.12
S.D. D_1—2b	12	13.8	10.7	.51	.71
Mean D_2—4b	12	68.08	74.00	1.57	2.32
S.D. D_2—4b	12	11.5	19.9	.52	.66
Mean Total 2b	47	67.38	70.11	1.58	2.39
S.D. Total 2b	47	12.4	15.2	.44	.72
Mean Total 4b	47	65.55	69.43	1.55	2.29
S.D. Total 4b	47	14.1	21.8	.41	.67

all sections were included. Therefore, the groups of pairs were considered to be equated, using as bases of pairing the marks in animal biology, average honor points in required premedical subjects, and scores on the Miller test. Three subgroups, as well as the group of twenty-seven, were also equated on the basis of the Sigma X reading test.

Table II gives a summary of the data for all students in the divisions, counting those not actually paired. It is evident that the inclusion of all students increases the number under consideration. If these divisions can be shown to be equated on the same bases as the divisions when only the paired students are counted, it will furnish a larger number from which to draw conclusions, and give us, therefore, more reliable data.

From calculations to determine the significance of the differences, there was found to be but one significant difference in the means, viz., that between the divisions of Section B in scores on the Miller test; and three significant differences between the standard deviations, viz., those between the divisions of Section D in scores on the Reading Test Sigma X and the total groups in scores on Sigma X. While the groups of students including those not paired were not so well equated as the groups of paired students, they were nearly enough equated to study in this experiment. Their relative achievements were later compared and used as bases from which conclusions may be drawn.

Plan of Instruction

In carrying on the work of the course all factors of instruction were kept as constant as was feasible, except this matter of the groupings of four's and two's. There were differences in instructors for the four sections and they were in different rooms during the laboratory periods, but each instructor had an equal number of groups of four's and two's, and those under each instructor were in the same room together. Besides laboratory work, which occupied twelve hours of work per week, all students had a one-hour recitation in osteology in Sections A, B, C, and D, all under one instructor; a one-half hour written test prepared by Dr. C. M. Jackson, head of the department, for all students alike; and a one-hour lecture to the whole group by Dr. Jackson. This made a total of fifteen hours work per week.

Measuring Instructional Outcomes

Considerable care was taken in devising measures of achievement to bring out any differences in the divisions which might exist, though this was recognized to be a very difficult task. It is admitted that there may be certain phases of achievement which were not measured, either because they have not yet been analyzed as entities, or cannot be easily measured. Nine half-hour objective tests of the uncompleted statement type were given during the quarter. Each of these tests contained fifty blanks, from which were omitted certain important key-words, to be filled in by the students, making a total of 450 objective items. At the end of the quarter a final written test covering about 250 items was given. One hundred and fifty of these items were repetitions of items in the nine half-hour tests in order to get data on reliability. There was also given in the final, a practical examination of 100 items on the work covered. This consisted of identification by the student of specimens, dissected in many different ways, the various particulars being indicated by tags. The students were given one minute on each item. Included in this practical test were eight specimens inclosed in boxes but cut off from the view of the students by curtains. The identifications were to be made only by means of the tactual sense, the student being requested to put his hand in the box and judge by the sense of touch. These tests were prepared by Dr. Jackson and given to all students alike. In addition to the test ratings, there were weekly estimates of each student's work by his laboratory instructor. These were expressed on a percentage basis. Quiz averages, the final written examination, the final practical examination, and the average laboratory estimates, all expressed in per cents, were averaged to give final numerical ratings. Letter ratings were then assigned as follows: below 70, F; 70–75, E; 75–79, D; 80–84, C; 85–89, B; above 90, A.

Tables III and IV give summaries of the data on final marks in letters for the whole group and for the pairs alone, respectively. Columns, "$Total_1$" and "$Total_2$" give the particular data for comparing the groups of two's and four's in all sections, respectively. From Table IV it is noticed that of the pairs working in two's, 11 received A's or B's, while of those working in four's, there were 13 A's or B's. Similarly, there were 16 receiving C, or lower, among

TABLE III

DISTRIBUTION OF LETTER MARKS IN ANATOMY, FALL QUARTER, 1926-27
(All Students)

MARK	SECTIONS										GRAND TOTAL
	A_1	A_2	B_1	B_2	C_1	C_2	D_1	D_2	TOTAL$_1$	TOTAL$_2$	
A	0	1	2	2	0	1	2	2	4	6	10
B	5	5	1	2	5	5	4	2	15	14	29
C	4	3	5	5	5	2	3	4	17	14	31
D	1	3	3	3	0	2	2	2	6	10	16
E	1	0	1	0	1	0	1	2	4	2	6
F	0	0	0	0	1	1	0	0	1	1	2
Total	11	12	12	12	12	11	12	12	47	47	94

TABLE IV

DISTRIBUTION OF LETTER MARKS IN ANATOMY, FALL QUARTER, 1926-27
(Pairs Only)

MARK	SECTIONS										GRAND TOTAL
	A_1	A_2	B_1	B_2	C_1	C_2	D_1	D_2	TOTAL$_1$	TOTAL$_2$	
A	0	0	0	1	0	0	2	2	2	3	5
B	4	4	0	2	3	2	2	2	9	10	19
C	3	3	4	2	1	1	2	2	10	8	18
D	0	1	2	2	0	2	1	0	3	5	8
E	1	0	1	0	1	0	0	1	3	1	4
F	0	0	0	0	0	0	0	0	0	0	0
Total	8	8	7	7	5	5	7	7	27	27	54

those working in two's; and 14 receiving C, or lower, among those working in four's. The differences here slightly favor those working in four's. For the whole number, as shown in Table III, there are 19 A's or B's among those working in two's, and 20 A's or B's among those working in four's. Similarly, there are 28 C's, or lower, among those working in two's and 27 C's, or lower, among those working in four's. Here also, the data slightly favor those working in four's.

Table V presents a summary of the mean achievement ratings in the objective tests, laboratory estimates, final practical tests, final written tests, and all of them combined, for the groups of

TABLE V

SUMMARY OF ACHIEVEMENT RATINGS (NUMERICAL) IN ANATOMY, FALL QUARTER, 1926-27
(Pairs Only)

	No. Pairs	Ratings				
		Mean 9 Objective	Mean Lab., 10 Weeks	Final Practical	Final Written	Mean of All
Mean A_1—2b.....	8	83.25	87.50	75.00	87.12	83.25
S.D. A_1—2b.....	8	7.50	2.20	11.30	7.30	6.40
Mean A_2—4b.....	8	83.25	87.62	78.63	85.00	83.75
S.D. A_2—4b.....	8	6.00	3.30	8.80	4.10	4.00
Mean B_1—2b.....	7	79.72	83.86	72.57	79.14	79.14
S.D. B_1—2b.....	7	4.80	3.30	8.50	7.90	4.30
Mean B_2—4b.....	7	82.14	85.29	75.57	87.86	82.71
S.D. B_2—4b.....	7	5.00	2.30	6.00	5.90	4.30
Mean C_1—2b.....	5	83.60	84.60	77.80	86.00	83.00
S.D. C_1—2b.....	5	3.00	6.70	6.80	8.30	5.30
Mean C_2—4b.....	5	82.80	84.00	76.20	82.80	81.80
S.D. C_2—4b.....	5	8.20	4.00	7.90	7.50	5.80
Mean D_1—2b.....	7	86.14	86.57	78.71	87.43	84.71
S.D. D_1—2b.....	7	5.00	4.60	3.60	7.60	5.70
Mean D_2—4b.....	7	86.14	86.00	79.00	87.00	84.57
S.D. D_2—4b.....	7	7.30	6.00	9.60	5.20	6.40
Mean $Total_1$—2b..	27	83.22	85.78	75.85	84.93	82.52
S.D. $Total_1$—2b..	27	5.90	4.50	9.50	8.50	5.90
Mean $Total_2$—4b..	27	83.63	86.00	77.48	85.85	83.33
S.D. $Total_2$—4b..	27	6.80	4.30	8.40	5.90	5.20

paired students. An examination of this table reveals no great differences. When differences were studied with regard to the question of significance, there was found to be a significant difference only in the mean scores on the final written examination in Section B. Counting all sections there is not one significant difference.

Table VI gives the summary of the data on achievement for all students (counting those not paired).

In the study of the achievement ratings of all students in the various sections and in all sections combined, there were found to be no statistically significant differences either in mean ratings or

TABLE VI

SUMMARY OF ACHIEVEMENT RATINGS (NUMERICAL) IN ANATOMY, FALL QUARTER, 1926-27
(All Students)

	Ratings				
	Mean 9 Objective	Mean Lab., 10 Weeks	Final Practical	Final Written	Mean of All
Mean A_1—2b	81.64	87.36	75.09	86.27	82.73
S.D. A_1—2b	7.20	2.20	9.70	7.60	5.80
Mean A_2—4b	82.83	87.67	78.67	85.67	83.75
S.D. A_2—4b	6.80	3.60	9.10	5.50	4.90
Mean B_1—2b	82.08	86.42	75.92	83.33	82.00
S.D. B_1—2b	6.80	3.90	9.40	9.00	5.90
Mean B_2—4b	82.33	86.33	73.58	87.92	82.67
S.D. B_2—4b	5.20	2.60	6.80	7.10	4.30
Mean C_1—2b	80.83	85.58	75.67	82.83	81.25
S.D. C_1—2b	7.70	5.20	11.60	14.30	9.00
Mean C_2—4b	83.54	85.83	74.09	84.09	82.09
S.D. C_2—4b	8.40	7.10	13.60	8.70	8.70
Mean D_1—2b	84.50	85.33	75.33	85.67	82.83
S.D. D_1—2b	5.40	4.60	9.80	7.80	6.30
Mean D_2—4b	82.50	83.67	75.25	82.50	81.08
S.D. D_2—4b	8.10	5.90	10.20	8.60	7.20
Mean $Total_1$—2b	82.28	86.15	75.64	84.49	82.19
S.D. $Total_1$—2b	6.90	4.30	10.20	10.20	7.00
Mean $Total_2$—4b	82.79	85.89	75.43	85.06	82.30
S.D. $Total_2$—4b	7.20	5.30	11.30	7.80	6.50

in standard deviations. In all the data given, there have been no differences revealed which show any advantage for either the plan of two students working on a cadaver, or of four students working on a cadaver. If there were advantages for either, the tests and estimates used in rating the students did not reveal them. Either there were actually no real differences in achievement, or there were certain phases of achievement that the very comprehensive measurements used did not consider. The matter of ability to manipulate dissection instruments and actually to dissect may be one of these phases. It is doubtful, however, if this phase is a primary

objective of anatomy. Even this would be included in laboratory estimates given by the instructors.

The Experiment during the Winter Quarter, 1926-27

For the winter quarter, the groups of students were shifted. Those working in two's during the fall quarter worked in four's during the winter quarter. Those working in four's during the fall quarter worked in two's, the winter quarter. Table VII is given to make comparisons of pairs of students in two's and four's in the various sections, A, B, C, and D, as they were constituted during the winter quarter. These preliminary ratings correspond with those of the fall quarter.

TABLE VII

Summary of Data for the Equation of Groups in Anatomy, Winter Quarter, 1926-27
(Pairs Only)

	No. Pairs	Miller Test	Sigma X Test	Av. H.P. Required	Animal Biology	Industry
Mean A_1—4b	6	72.80	58.17	1.39	2.39	2.00
S.D. A_1—4b	6	9.80	15.90	.23	.52	.41
Mean A_2—2b	6	71.50	70.00	1.40	2.28	2.17
S.D. A_2—2b	6	8.70	21.90	.24	.52	.47
Mean B_1—4b	2	67.50	58.00	1.43	2.75	2.00
S.D. B_1—4b	2	.50	2.00	.24	.25	0.00
Mean B_2—2b	2	66.50	60.50	1.43	2.50	1.75
S.D. B_2—2b	2	2.50	8.50	.24	.50	.25
Mean C_1—4b	3	66.70	73.30	1.76	2.33	1.33
S.D. C_1—4b	3	8.50	16.90	.17	.27	.23
Mean C_2—2b	3	68.30	77.00	1.75	2.00	2.17
S.D. C_2—2b	3	8.00	5.70	.14	.72	.62
Mean D_1—4b	3	64.00	72.70	1.88	1.95	1.83
S.D. D_1—4b	3	7.10	9.20	.47	.52	.24
Mean D_2—2b	3	65.70	71.30	1.89	2.06	1.83
S.D. D_2—2b	3	9.90	19.90	.54	.44	.85
Mean All_1—4b	14	68.90	64.50	1.58	2.33	1.82
S.D. All_1—4b	14	9.00	15.50	.38	.51	.40
Mean All_2—2b	14	68.90	70.40	1.58	2.20	2.11
S.D. All_2—2b	14	8.60	18.20	.38	.58	.60

But fourteen of the original twenty-seven pairs remained in all sections. Counting these fourteen pairs there were no statistically significant differences in any of the criteria of equation either in means, standard deviations, or coefficients of variability.[4]

TABLE VIII

SUMMARY OF DATA FOR THE EQUATION OF GROUPS IN ANATOMY, WINTER QUARTER, 1926-27

(All Students)

	No. Students	Miller Test	Sigma X Test	Av. H.P. Required	Animal Biology
Mean A_1—4b	8	70.50	63.12	1.37	2.51
S.D. A_1—4b	8	12.20	17.70	.20	.55
Mean A_2—2b	12	68.92	70.67	1.67	2.04
S.D. A_2—2b	12	13.70	22.90	.43	.61
Mean B_1—4b	8	72.38	70.75	1.62	2.50
S.D. B_1—4b	8	13.80	15.00	.54	.56
Mean B_2—2b	12	57.42	56.50	1.41	2.64
S.D. B_2—2b	12	13.30	20.20	.28	.55
Mean C_1—4b	8	73.12	81.50	1.62	2.37
S.D. C_1—4b	8	9.50	14.70	.32	.49
Mean C_2—2b	10	68.50	75.10	1.58	2.07
S.D. C_2—2b	10	15.20	18.90	.31	.68
Mean D_1—4b	4	59.00	65.75	1.44	2.13
S.D. D_1—4b	4	10.60	14.40	.27	.55
Mean D_2—2b	12	68.08	74.00	1.57	2.32
S.D. D_2—2b	12	11.50	19.90	.52	.66
Mean All_1 4b	28	70.10	70.93	1.53	2.41
S.D. All_1 4b	28	12.70	17.30	.39	.55
Mean All_2 2b	46	65.61	69.02	1.55	2.28
S.D. All_2 2b	46	14.30	21.80	.41	.65

Table VIII gives a summary of all the students in the various sections who were working in two's and four's. A study of significances showed that there were no statistically significant differences between the means and variabilities of the total groups of two's and four's in all sections. There were then (1) the paired students

[4] Karl Pearson, *Tables for Statisticians and Biometricians*, p. 22.

in all sections forming two groups which were comparable in the criteria given and (2) the total numbers of students in all sections forming two groups, which were also comparable one with another except in number of students in these groups. These were twenty-eight and forty-six, respectively, in four's and two's. Considering the groups of pairs under four instructors, the two members of a pair were always under the same instructor. Considering the groups of all students (twenty-eight and forty-six), the students were well distributed among the four instructors. It will be remem-

TABLE IX

SUMMARY OF ACHIEVEMENT OF GROUPS IN ANATOMY
WINTER QUARTER, 1926-27
(Pairs Only)

	No. Students	Ratings				
		Quiz Average	Laboratory Average	Final Practical	Final Written	Mean of All
Mean A_1—4b.....	6	87.67	90.00	84.00	85.50	86.67
S.D. A_1—4b.....	6	2.3	1.8	6.1	3.6	2.7
Mean A_2—2b.....	6	81.17	88.50	74.33	81.33	80.83
S.D. A_2—2b.....	6	7.7	2.1	10.0	6.3	6.0
Mean B_1—4b.....	2	82.50	87.50	64.00	72.00	77.00
S.D. B_1—4b.....	2	.5	1.5	9.0	9.0	5.0
Mean B_2—2b.....	2	77.00	87.50	77.00	83.50	81.50
S.D. B_2—2b.....	2	6.0	7.5	13.0	6.5	6.6
Mean C_1—4b.....	3	80.33	87.67	81.00	79.50	82.33
S.D. C_1—4b.....	3	2.1	3.4	1.4	5.2	.9
Mean C_2—2b.....	3	87.33	86.67	84.33	85.67	86.33
S.D. C_2—2b.....	3	3.0	2.5	2.1	6.1	2.5
Mean D_1—4b.....	3	80.67	86.33	78.00	77.67	81.00
S.D. D_1—4b.....	3	8.3	2.6	14.0	10.4	8.8
Mean D_2—2b.....	3	81.33	87.33	73.00	89.50	80.67
S.D. D_2—2b.....	3	6.5	4.6	15.6	.05	9.7
Mean $Total_1$—4b..	14	83.86	88.36	79.21	80.64	83.14
S.D. $Total_1$—4b..	14	5.4	3.2	10.6	8.3	6.0
Mean $Total_2$—2b..	14	81.93	87.71	76.57	82.29	82.14
S.D. $Total_2$—2b..	14	7.2	4.1	11.8	6.9	7.2

bered that each instructor had approximately half of his students working in two's and half in four's.

The achievements of students in the winter quarter were tabulated similarly to those in the fall quarter and were secured from similar sets of quizzes, laboratory estimates, and final examinations. Table IX presents the data for the groups of paired students. A study of the significances of differences in these achievement ratings for the group of pairs shows no significant differences in either means or standard deviations for any of the ratings. The small number of students under individual instructors made it inadvisable to consider the separate divisions. Only the pairs in all

TABLE X

SUMMARY OF ACHIEVEMENT OF GROUPS IN ANATOMY
WINTER QUARTER, 1926-27
(All Students)

	No. Students	Ratings				
		Quiz Average	Laboratory Average	Final Practical	Final Written	Mean of All
Mean A_1—4b.....	8	84.75	89.12	81.62	81.50	84.25
S.D. A_1—4b.....	8	5.6	2.3	6.7	7.6	4.8
Mean A_2—2b.....	12	82.75	89.00	78.92	81.67	83.25
S.D. A_2—2b.....	12	7.4	2.6	10.4	7.7	6.2
Mean B_1—4b.....	8	85.12	89.12	75.37	77.63	82.12
S.D. B_1—4b.....	8	5.6	3.4	13.7	10.0	7.8
Mean B_2—2b.....	12	82.33	84.43	76.25	78.75	80.75
S.D. B_2—2b.....	12	5.3	4.9	8.2	9.7	5.4
Mean C_1—4b.....	8	80.87	88.62	81.37	85.37	84.75
S.D. C_1—4b.....	8	4.2	2.5	4.1	6.1	3.0
Mean C_2—2b.....	10	83.10	87.70	82.50	82.90	83.80
S.D. C_2—2b.....	10	7.2	5.1	6.5	5.2	5.4
Mean D_1—4b.....	4	82.25	86.50	78.25	76.75	81.00
S.D. D_1—4b.....	4	7.2	3.6	12.1	5.3	7.6
Mean D_2—2b.....	12	82.17	87.08	77.92	83.33	82.75
S.D. D_2—2b.....	12	6.5	4.8	12.2	7.3	7.5
Mean $Total_1$—4b..	28	83.29	86.61	79.29	80.82	83.32
S.D. $Total_1$—4b..	28	5.9	3.0	10.0	8.8	6.1
Mean $Total_2$—2b..	46	82.57	87.13	78.74	81.61	82.59
S.D. $Total_2$—2b..	46	6.6	4.7	9.9	8.1	6.3

sections were considered. What differences there were in the groups of pairs favored the four's except in the final written tests. This held also counting all students.

Table X gives achievement ratings for all students in the groups. Counting all students, there was a significant difference in the standard deviations of the laboratory estimates, those working in two's having a significantly greater variability. The difference in means in this rating was nearly great enough to be really significant.

TABLE XI

DISTRIBUTION OF LETTER MARKS IN ANATOMY, WINTER QUARTER, 1926-27*
(All Students)

MARK	SECTIONS											
	A_1	A_2	B_1	B_2	C_1	C_2	D_1	D_2	ALL$_1$	ALL$_2$	ALL$_1$	ALL$_2$
											%	%
A	2	3	2	1	1	1	1	3	6	8	21.4	17.4
B	2	3	1	3	2	5	0	3	5	14	17.8	30.4
C	2	2	3	4	5	2	2	3	12	11	42.8	23.9
D	2	3	1	3	0	1	0	1	3	8	10.7	17.4
E	0	1	1	1	0	1	1	1	2	4	7.1	8.7
F	0	0	0	0	0	0	0	1	0	1	0	2.2
Total	8	12	8	12	8	10	4	12	28	46	99.8	100.0

TABLE XII

DISTRIBUTION OF LETTER MARKS IN ANATOMY, WINTER QUARTER, 1926-27*
(Pairs Only)

MARK	SECTIONS									
	A_1	A_2	B_1	B_2	C_1	C_2	D_1	D_2	ALL$_1$	ALL$_2$
A	2	0	0	1	0	0	1	1	3	2
B	2	2	0	0	0	2	0	0	2	4
C	2	1	1	0	3	1	1	1	7	3
D	0	2	1	1	0	0	0	0	1	3
E	0	1	0	0	0	0	1	0	1	1
F	0	0	0	0	0	0	0	1	0	1
Total	6	6	2	2	3	3	3	3	14	14

* Divisions 1 ($_1$) worked together in groups of four's; Division 2 ($_2$) worked together in groups of two's.

The mean was higher for those working in two's. This rating is a subjective estimate, and the differences in means and standard deviations were not great, even though the latter were significant. Objective ratings, however, did not show significant differences. What differences there were, favored those working in two's in some cases; those in four's in other cases.

Table XI shows the distribution of letter marks among students working in two's and four's, respectively, counting all students. Table XII gives similar data for the paired students alone. Both show more A's among those in four's; more B's among those in two's; more C's among those in four's; more D's among those in two's; and more F's among those in two's. Those getting E's were about equal in both. There is no distinct superiority for either grouping.

Individual Students under the Two Plans

The accounts of the experiment for the fall and winter quarters, respectively, have been based on the comparisons of equated groups of students. These comparisons involved the question of equating groups and of pairing students for the equated groups. By carrying on the experiment for two successive quarters, it became possible, however, to compare a given student's achievement working under one set of laboratory conditions (i.e., working in two's) with his own achievement under the contrasted set of conditions (i.e., working in four's). His achievement would not, therefore, be compared with that of a paired companion but with that of himself when working under two contrasted conditions. In order to obviate the possible effect that one method might have on the succeeding method, both possibilities have been met. For example, one set of students worked first in two's, then in four's, during the fall and winter quarters, respectively. The other set of students worked first in four's, then in two's, for the same quarters, respectively. Table XIII presents a summary of the relative ratings for the students first working in two's; Table XIV, for those first working in four's.

A study of Tables XIII and XIV indicates apparent advantages for some students when working in two's and for other students when working in four's. A slightly greater advantage is shown

TABLE XIII

SUMMARY OF RELATIVE RATINGS OF STUDENTS IN ANATOMY, 1926-27

	QUARTER	NUMBER	PER CENT
Same letter marks		16	57.1
Higher marks in four's	Winter	5	17.8
Higher marks in two's	Fall	7	25.1
Total		28	100.0

7.3 per cent more had higher marks in two's.

No displacement of more than one position, positions being A, B, C, D, E, and F.

TABLE XIV

SUMMARY OF RELATIVE RATINGS OF STUDENTS IN ANATOMY, 1926-27

	QUARTER	NUMBER	PER CENT
Same letter marks		21	46.0
Higher marks in four's	Fall	13	28.0
Higher marks in two's	Winter	12	26.0
Total		46	100.0

2.0 per cent more had higher marks in four's.

Two students displaced two positions. All other displacements—one position, positions being A, B, C, D, E, F.

for those working in two's when this occurred the first quarter. Similarly, the advantage lay with the four's when this occurred the first quarter. Whether or not there would be a difference in the quarters' marks, were they working in the same groups two succeeding quarters, is a question.

It was thought worth while to compare the achievements, not only as given in final letter marks, but in the various phases of achievement entering into the final letter mark. Therefore, quintile ratings were prepared, dividing the total group of 82 students, who carried through the work for both quarters, into five equal groups on the basis of each measured phase of achievement. These show whether a student placed in the lowest 20 per cent, the second, third, fourth, or highest 20 per cent of achievement. The phases of achievement considered were (1) the final written examination; (2) the final practical examination; (3) the average quiz ratings; and (4) the laboratory estimates. Tables XV to XXII, inclusive,

present summaries for students working first in two's and students working first in four's, with comparisons of these students under the two contrasted plans, respectively.

A study of the tables shows a slight tendency for the grouping of two's to do better on the average than the grouping of four's.

TABLE XV

SUMMARY OF RELATIVE RATINGS OF STUDENTS IN FINAL WRITTEN TEST*

	QUARTER	NUMBER	PER CENT
Same quintiles		8	28.6
Higher quintiles in four's	Winter	9	32.1
Higher quintiles in two's	Fall	11	39.3
Total		28	100.0

7.2 per cent more in higher quintiles in two's.

* In the test in Anatomy, 1926-27, for which data are given in Tables XV through XXII the students were *classed* according to percentile rank: 1 being the lowest, 5, the highest.

TABLE XVI

SUMMARY OF RELATIVE RATINGS OF STUDENTS IN FINAL WRITTEN TEST

	QUARTER	NUMBER	PER CENT
Same quintiles		13	28.2
Higher quintiles in two's	Winter	17	36.9
Higher quintiles in four's	Fall	16	34.9
Total		46	100.0

2.0 per cent more in higher quintiles in two's.

TABLE XVII

SUMMARY OF RELATIVE RATINGS OF STUDENTS IN FINAL PRACTICAL TEST

	QUARTER	NUMBER	PER CENT
Same quintiles		15	53.5
Higher quintiles in four's	Winter	6	21.4
Higher quintiles in two's	Fall	7	25.1
Total		28	100.0

3.7 per cent more in higher quintiles in two's.

TABLE XVIII

SUMMARY OF RELATIVE RATINGS OF STUDENTS IN FINAL PRACTICAL TEST

	QUARTER	NUMBER	PER CENT
Same quintiles		19	41.3
Higher quintiles in four's	Fall	13	28.3
Higher quintiles in two's	Winter	14	30.4
Total		46	100.0

2.1 per cent more in higher quintiles in two's.

TABLE XIX

SUMMARY OF RELATIVE RATINGS OF STUDENTS IN QUIZ AVERAGES

	QUARTER	NUMBER	PER CENT
Same quintiles		13	46.4
Higher quintiles in four's	Winter	7	25.0
Higher quintiles in two's	Fall	8	28.6
Total		28	100.0

3.6 per cent more in higher quintiles in two's.

TABLE XX

SUMMARY OF RELATIVE RATINGS OF STUDENTS IN QUIZ AVERAGES

	QUARTER	NUMBER	PER CENT
Same quintiles		17	36.9
Higher quintiles in four's	Fall	16	34.7
Higher quintiles in two's	Winter	13	28.4
Total		46	100.0

6.3 per cent more in higher quintiles in four's.

TABLE XXI

SUMMARY OF RELATIVE RATINGS OF STUDENTS IN LABORATORY WORK

	QUARTER	NUMBER	PER CENT
Same quintiles		12	42.8
Higher quintiles in four's	Winter	5	17.8
Higher quintiles in two's	Fall	11	39.4
Total		28	100.0

21.6 per cent more in higher quintiles in two's.

TABLE XXII

SUMMARY OF RELATIVE RATINGS OF STUDENTS IN LABORATORY WORK

	QUARTER	NUMBER	PER CENT
Same quintiles		17	36.9
Higher quintiles in four's	Fall	15	32.6
Higher quintiles in two's	Winter	14	30.5
Total		46	100.0

2.1 per cent more in higher quintiles in four's.

Certain individuals did better under each condition, but it could not be proved that this was because of the particular condition. For example, Table XV indicates that nine did better in four's; eleven did better in two's. The final practical favors the two's slightly also, whether or not the grouping in two's comes first. Quiz averages and laboratory estimates favor each of the two groups (in two's and in four's) when it comes first but the greater advantage in quiz averages is with the four's and in laboratory estimates with the two's.

TABLE XXIII

SUMMARIES OF TABLES XIII TO XXII (INCLUSIVE)*

FINAL LETTER MARKS

	NUMBER	PER CENT
Same letter marks	37	50.0
Higher marks in four's	18	24.3
Higher marks in two's	19	25.7
Total	74	100.0

1.4 per cent more in higher quintiles in two's.

FINAL WRITTEN TEST

	NUMBER	PER CENT
Same quintiles	21	28.3
Higher quintiles in four's	25	33.8
Higher quintiles in two's	28	37.9
Total	74	100.0

4.1 per cent in higher quintiles in two's.

* Without reference to the quarter.

TABLE XXIII—*Continued*

SUMMARIES OF TABLES XIII TO XXII (INCLUSIVE)

FINAL PRACTICAL TEST

	NUMBER	PER CENT
Same quintiles	34	46.0
Higher quintiles in four's	19	25.6
Higher quintiles in two's	21	28.4
Total	74	100.0

2.8 per cent more in higher quintiles in two's.

QUIZ AVERAGES

	NUMBER	PER CENT
Same quintiles	30	40.6
Higher quintiles in four's	23	31.1
Higher quintiles in two's	21	28.3
Total	74	100.0

2.8 per cent more in higher quintiles in four's.

LABORATORY ESTIMATES

	NUMBER	PER CENT
Same quintiles	29	39.2
Higher quintiles in four's	20	27.0
Higher quintiles in two's	25	33.8
Total	74	100.0

6.8 per cent more in higher quintiles in two's.

Tables XIII to XXII, inclusive, contain data for the single groups divided according to the quarters, fall and winter, respectively. Table XXIII, however, presents summaries of the data included in Tables XIII to XXII but combining the students for both quarters. The contrast was made in Table XXIII between the groupings of two's and four's without reference to the quarter in which either method was followed. A slight advantage for the grouping in two's is shown in all ratings except the quiz averages. Always, however, there are certain individuals who have higher ratings in four's.

The data on letter marks in Table XXIII summarizes the data on the change in letter marks of individuals under the two methods. Another method of comparing the achievements in the two contrasted methods of two's and four's, respectively, was used in tabulating the letter marks received by all students who took both methods to determine the number receiving each letter mark under the contrasted methods. This tabulation shows merely the number of letter marks received under the two methods, without reference to whether or not there was a change in such marks for given individuals. Table XXIV presents the result of this tabulation.

TABLE XXIV

Distribution of Letter Marks in Anatomy, 1926-27, Counting Only Those Who Reversed

Marks	Two on a Body			Four on a Body		
	Quarter			Quarter		
	First	Second	Both	First	Second	Both
A	3	8	11	6	6	12
B	9	14	23	14	5	19
C	13	11	24	14	12	26
D	3	8	11	10	3	13
E	0	4	4	2	2	4
F	0	1	1	0	0	0
Total	28	46	74	46	28	74

One more student received an A under the plan of four's; four more students received B's under the plan of two's; two more students received C's under the plan of four's; the number of D's were the same under the two plans; and one student received a D under the plan of two's, while there were no D's received by students under the plan of four's. There is no evidence from these data to show any decided superiority for either plan.

Ratings of the Anatomy Class of 1925-26 for Comparison

The comparison made between ratings extending over two quarters of work, which has been discussed in the preceding pages, introduced the question of quarter variations in marks in anatomy

when the students used the same method of grouping to a cadaver for both quarters. It was noticed that marks from fall quarter to winter quarter varied considerably when students were transferred from groups of two's to groups of four's and vice versa. Would there be as much variation between marks for two succeeding quarters if students worked in two's both quarters? During the school year of 1925-26, students in anatomy worked in two's both quarters. It seemed worth while to study the available ratings, to notice any variations existing from quarter to quarter. While the tests were not the same as those used in 1926-27, they were similar. Final marks were computed in a similar manner. For this comparison eighty-two students carrying work for the two successive quarters of 1925-26 were selected. These students were made comparable with those of the eighty-two students studied in the 1926-27 classes by selecting students with the same letter ratings in anatomy at the end of the fall quarter as those of the 1926-27 classes. Thus, twenty-eight students from the classes of 1925-26 were selected to be compared with the twenty-eight from the 1926-27 classes (Tables XIII to XXIII); and forty-six, from the classes of 1925-26 to compare with forty-six from the 1926-27 classes (Tables XIII to XXIII).

Tables were constructed for direct comparison, respectively, with each of Tables XIII to XXIII, inclusive. Space does not permit the inclusion of these tables here, but it may be stated that there were apparent, from quarter to quarter, even wider variations in marks given to individuals. One difference was that there was noticeable a tendency during 1925-26 for higher marks during the second quarter, while for 1926-27 there was no such noticeable tendency.

There seems to be no way to determine whether the variations of the ratings in the 1926-27 class from quarter to quarter are due to variations due to a grouping in two's and four's or to other factors dependent upon natural changes in the quarters themselves. The only evidence seems to indicate that, as far as the ratings available were concerned, a settled technique in the laboratory for both quarters seemed to cause higher ratings the second quarter, while a change from groupings of two's to four's and vice versa seemed to cause less variation. Correlations between fall and

winter quarter ratings for the years 1925-26 and 1926-27, respectively, lead to the same conclusion. For 1925-26 the correlation between final marks in anatomy for fall and winter, respectively, was .65 ± .04. For the same quarters, 1926-27, the coefficient was .92 ± .01.

Correlations

The question of the validity of the criteria used in pairing students for the equated groups studied in the experiment naturally arises. It was noticed that when students were paired on the basis of one criterion that they did not always pair on the basis of some other criterion, though usually, the mean ratings of groups of such students approximated equality in any set of ratings used. It would be desirable to have such a criterion of equality that correlations between ratings in the criterion and in the achievement ratings would approach unity. It was only at the conclusion of

TABLE XXV

Correlations between Final Ratings in Anatomy and Various Criteria*

Other Rating	Quarter	Year	Coefficient	P.E.	No. of Cases
Miller test	Fall	1926-27	.24	.07	92
Miller test	Winter	1926-27	.24	.07	88
Reading Test Sigma X	Fall	1926-27	.13	.07	93
Reading Test Sigma X	Winter	1926-27	.27	.06	87
Average required honor points	Fall	1926-27	.18	.07	92
Average required honor points	Winter	1926-27	.32	.06	87
Animal biology	Fall	1926-27	.32	.06	93
Animal biology	Winter	1926-27	.46	.06	88
Average total honor points	Fall	1926-27	.25	.07	92
Average total honor points	Winter	1926-27	.30	.06	88
Winter 1926-27	Fall	1926-27	.92	.01	88
Winter 1925-26	Fall	1925-26	.65	.04	81
Miller test (objective)	Fall	1926-27	.26	.06	84
Animal biology and Miller test (multiple)	Fall	1926-27	.36	.06	88
Animal biology and Miller test (multiple)	Winter	1926-27	.48	.06	82

The coefficient between animal biology and the Miller test was .22±.07 for ninety-four cases.

* Otis Correlation Chart used in Tables XXV and XXVI.

this experiment that steps were taken to establish the validity of the criteria used in pairing. This was done by finding the coefficient of correlation between the final achievement ratings and each of the criteria. Table XXV lists these correlations. The correlations tell their own story. There are comparatively low correlations between achievement in anatomy and the various criteria. No previous record of the student seems to be very useful in forecasting his rank in achievement in anatomy. In other words present data for pairing students for an experiment in anatomy such as this are inadequate. The high correlation, however, between achievement in the fall and winter, respectively (especially high for the year 1926-27), gives further evidence of the probable truth that working in two's or four's is immaterial in respect to the ratings obtained.

The correlations in Table XXVI are given for reliability and validity considerations.

TABLE XXVI

Correlations between Other Tests

	Coefficient	P.E.	No. of Cases
Final practical-visual and final practical-tactual (fall quarter)	.32	.06	90
Test 1 objective and its repetition about 11 weeks later	.35	.06	87
Test 3 objective and its repetition about 7 weeks later	.51	.05	90
Test 10 objective and its repetition about 1 week later	.66	.04	88
Final practical-visual and final practical-tactual (winter quarter)	.43	.06	90

Discussion of Findings

This experiment was undertaken, primarily, to discover if there were benefits obtained by furnishing a cadaver for every two students not obtained by furnishing only one for every four students. No conclusive evidence has been procured which shows that it is essential that one cadaver be furnished for every two students. While certain individual students have higher ratings when sharing a cadaver with but one other student, practically an equal number

have higher ratings when sharing a cadaver with three other students. There is evidence indicating that the number of students to a cadaver is not an important determinant of achievement. Paired groups of students show no distinct superiority by either method. No general tendency is evident when achievements of given individuals under the two methods are studied. In extensive written examinations, in quizzes, in identification of anatomical structures, in general work in the laboratory, the students who work in four's seem, on the average, to do as well as students who work in two's. When data regarding groups of paired students, all students in comparable groups, individual students paired with one another, or given individuals working under the two compared conditions are examined and studied, no evidence shows a consistent superiority of either procedure over the other.

The tests used have covered the informational content thoroughly, more than 1,000 separate items being included in all written tests or examinations. As anatomy is primarily an informational subject, these tests and examinations should furnish valid evidence. We have given no statistical evidence of validity of the tests as such, but this seems unnecessary. An examination of the tests is sufficient to convince one of the fact that the items are strictly anatomical. One who had not worked and studied in anatomy would be unable to cope with them at all.

Not much statistical data have been given bearing on reliability. It has been almost impossible to get these data because of the nature of the instructional material and the procedures of instruction. The material is so specific and definite and the learning so continuous from day to day in the specific field, that repetitions of the same test would be apt to give varying results. Three correlations have been given for certain first-term tests and their repetitions in the final examinations in Table XXVI. The correlations have increased as the time between the repetitions has decreased. It would be logical to believe that if a repetition followed at a short-time interval, the correlation would be very high.

Whether students working in two's benefit in some manner not shown by the measurements made, is a question. A questionnaire in which the instructors recorded their reasons for preferences for

students working in two's gives as reasons for such preferences the following:

1. It enables each student to do more individual dissection.
2. It affords more opportunity for development of initiative and self-confidence.
3. It gives first-hand information.
4. It is difficult to get four who work well together.
5. The weak student is compelled to do the work for himself.
6. Four is too much of a crowd.

These citations introduce the question of the aims and objectives of the course. What is the function of the anatomy course? Possibly, if skill in dissection is an important objective, opportunity for more practice in dissection offered by groupings in two's would be desirable. In order to verify this measurements should be made in this particular matter. It cannot be accepted without more proof that the greater opportunity offered in dissection by groupings of two's over that offered by groupings of four's would actually, of itself, produce more such skill. It seems reasonable that, as opportunity for acquiring skill in dissection is given later in the curriculum, the informational function of the course is its major function. Initiative and self-confidence are certainly desirable, but there is no proof that these are obtained from groupings of two's over that of four's. It might even be argued that meeting the competition of three others successfully would stimulate these characteristics more fully. As to first-hand information the measurements all indicate that those working in four's furnish as complete information. There may still be a question as to whether they will retain this information as long.

There is no evidence available to substantiate the other contentions noted. Number 4 (it is difficult to get four who work well together) may be true, but evidently the difficulty is not insurmountable, merely an inconvenience. Number 5 (the weak student is compelled to do the work for himself) does not prove that, possibly, a weak student might gain more real knowledge of anatomy by being allowed to work with stronger students. Number 6 (four is too much of a crowd) represents merely a feeling that is not substantiated by any of the facts brought out. There may be preferences which might receive recognition if no other reasons

contraindicated. Apparently, no serious damage would result by denying such privileges in this matter.

Questionnaire returns from the students give a large majority preference for the grouping in two's: seventy-five for, fourteen against, with two having no preference. The fourteen voting in favor of the grouping in four's received no higher marks when working in four's. Therefore, it seems an unprejudiced vote. The advantages summarized for each of the groupings are interesting. The arguments for working in two's are similar to those given by the instructors. The advantages summarized in favor of the groupings in four's seem fairly well concentrated on the saving of time by division of labor and the advantages of more opportunities for discussion.

How much student and faculty opinion is influenced by the prevailing procedures, which have been followed for some time, is unknown. It is evident that the facts have not substantiated the opinions expressed in favor of the groupings in two's. It seems reasonable to believe that, if there were such overwhelming advantages of the groupings in two's, it would have shown up somewhere in all the measurements made.

It is worth while to note that, of the various methods used in making comparisons, there is no disagreement shown in any one method. When the paired students are treated as individuals or as groups, when all students are considered in groups, or when individuals are considered as working under the different conditions (two's and four's)—the data all point to similar conclusions. Therefore, the evidence here points to the validity of making the comparisons by any one of the methods used.

However, there is no conclusive evidence given which proves definitely under which of the conditions (two's or four's) any given individual will make the greater achievement—only, neither condition is preeminently superior to the other for all members of a group.

A Comparison of Sections A and Sections B, in Advanced Courses in the Medical School

Are there any differences between Sections A and B in subsequent courses in the Medical School? Data were obtained for courses in neurology, histology, Physiology 103, and Physiology 104.

To answer this question tables were constructed giving the means, standard deviations, and coefficients of variability for the sections for each of the courses, respectively; and statistical comparisons for significance were made, using the difference ÷ the standard error of the difference = 3 as a criterion of significance. The tables are omitted here for lack of space but in not a single instance did this ratio exceed 2.4, and in but this one case did it exceed 1.6. In twenty-one times out of twenty-eight, it was less than 1. In other words, the differences were not significant. The correspondence between the groups in the different ratings was the striking feature. Table XXVII shows the number of times each section exceeded the other section in the various courses.

TABLE XXVII

Number of Times Sections A Exceed Sections B in the Mean Ratings in All Courses and Vice Versa

	Neurology	Histology	Physiology 103	Physiology 104	Total
Sections A.....	5	2	4	3	14
Sections B.....	2	5	3	4	14

An examination of Table XXVII makes it evident that Sections A and B exceeded each other an equal number of times, viz., 14. Considering only final averages in each course, however, Sections A exceeded Sections B three times, while Sections B exceeded Sections A but once. The ratings showing the largest differences in the four courses between Sections A and B, were found in the final practical, three times, and in the laboratory once. Two of these differences favor Sections A, and two, Sections B.

In examining for differences in variabilities of the groups for all ratings given, Sections A were more variable six times, Sections B were more variable seventeen times and coefficients of variability were the same five times. The differences, however, were small in every case. The greatest differences were in the final practical test. The differences in both mean ratings and in coefficients of variation seem to follow the laws of chance. In other words, neither of the groups of sections shows superiority over the other. As groups, then, the original groupings of students in anatomy in two's and four's, respectively, seem to have had no effect in producing differences in ratings in more advanced courses in the Medical School.

CHAPTER IV

LABORATORY VERSUS LIBRARY IN THE TEACHING OF HUMAN PHYSIOLOGY

The experiment to determine the effort of replacing a part of the laboratory work by library work was performed during the fall quarter, 1926-27, at the University of Minnesota in the Department of Human Physiology of the Medical School, to determine the effect of reducing the number of hours per week given to laboratory work in human physiology. Previously, seven and a half hours a week had been given to laboratory work. It was thought wise to try out with one section a schedule of five hours of laboratory work with two and a half hours spent on assigned library work. Before making this change permanent in the schedule, it was thought best to make an attempt to evaluate the changed procedures by comparing the results with those obtained under the old plan. It is obvious that if students could gain as much or more value from less formal laboratory work, economy would result. Such a scheme would necessitate less instructional cost in instructors, laboratory materials required, and possibly less laboratory space. Certainly, laboratory instruction should not be given unless it definitely shows certain results gained, not obtainable by other less costly procedures.

The regular work of the course consisted of five lectures a week, two quizzes a week, and seven and one-half hours of laboratory work. For this experiment two equated sections of students were used: Section A was known as the control group maintaining the old schedule; Section B was given the changed schedule, viz., five hours of laboratory work and two and one-half hours of library work (the latter to be a substitute for the time taken from laboratory work). The plan of the experiment was to keep all factors of learning over which the school had control constant in the two sections, the only difference being the substitution of the two and one-half hours of library work for an equal amount of laboratory work. Identical lectures were to be given to both sections by Dr. F. H. Scott; the quizzes were to be given to the sections by Dr. Scott and Dr. E. P. Lyon, the two to reverse sections at the mid-quarter so that each would spend half the time with each section;

the laboratory work of both sections was to be supervised by the same two instructors, Mr. Loucks and Mr. Steggerda, together; the library work was to be directed by Instructor King, the work bearing directly on the work in physiology. The plan adopted for the library work was as follows: A list of topics was made up, from which list each student in Section B drew by lot for his assignment. Each student was to prepare a talk on his topic to be given before the class.

In order to equate Sections A and B, two psychological tests were given the first day of the quarter to all members of the sophomore class. The tests chosen were the Miller mental ability test for graduate students and the Sigma X reading test. The number of honor points gained by each student during the freshman year, based on a total of thirty-nine credits in anatomy, histology, and physiological chemistry, were also computed. For Sections A and B students were selected who were very nearly comparable in the Miller test and in the number of freshman honor points gained.

TABLE XXVIII

Summary of Group Comparisons in Physiology

	Average Honor Points Required	Ratings		
		Miller	Fresh. H. P.	Sigma X
Section A* Mean	1.56	69.9	59.9	73.2
S.D.	.427	11.3	24.9	17.5
Section B† Mean	1.66	67.8	51.9	74.7
S.D.	.468	12.1	25.8	21.1

* Seven and one-half hours of laboratory work.

† Five hours of laboratory work and two and one-half hours of library work.

These two criteria were judged to be of the most value in pairing the students, though no data were available to verify this judgment. It was found possible by this method of pairing to get but twenty-two students in Section A paired with students in Section B. Later this was reduced to twenty by changes in enrolment. The final roster of Sections A and B comprised thirty-two students in each section, twenty of them being paired, the remaining twelve being selected at random. As further means of evaluating the equation

of the sections, the average honor points of the students in the required subjects of their premedical courses were obtained. Comparisons were made, not only of the paired students, but of all the students, considering them as groups.

Table XXVIII gives the summary of the data as to the constitution of the groups, counting all students including those paired in the four criteria mentioned. The points of comparison to be noted are the means and standard deviations of the groups. It may be noticed that the sections were not identical in any of the four criteria used. Originally, each contained thirty-two students. Incomplete records for the quarter eliminated two from Section B, so the group preliminary ratings were revised to parallel the final subject-matter ratings to be given further on. Section A thus contained thirty-two students with complete records, and Section B, but thirty. Section A appeared superior in the Miller test and in total honor points in freshman subjects; while Section B appeared superior in average honor points in required premedical subjects and in the Reading Test Sigma X.

When, however, these figures were subjected to the statistical treatment of probable errors due to the inadequacy of the sampling on account of the small number of students used,[1] there appeared to be no significant differences in the groups in any of the criteria. Table XXIX gives these data.

TABLE XXIX

RELIABILITIES OF DIFFERENCES BETWEEN SECTIONS A AND B IN PHYSIOLOGY*

	Freshman Honor Points	Miller Test	Sigma X Test	Average Honor Points
Difference in means ÷ P.E. of difference	1.8‡	1.0‡	4§	1.2§
Difference in S.D.'s ÷ P.E. of difference	.3§	.6§	1.5§	.7§

* In this and succeeding tables ‡ indicates A higher; §, B higher. There are no significant differences.

From these data the groups may be considered as equated. The groups of pairs in the two sections were also considered as groups. Table XXX presents these data. The ratings in industry are the

[1] Edward L. Thorndike, *loc. cit.*

TABLE XXX

SUMMARY OF DATA FOR THE EQUATIONS OF GROUPS, SECTIONS A AND B IN PHYSIOLOGY
(Pairs Only)

		Av. H. P. Required	Miller Test	Freshman Honor Points	Sigma X Test	Industry
Section A	Mean.....	1.53	70.1	54.6	67.7	1.97
	S.D.....	.44	8.8	22.5	14.0	.60
Section B	Mean.....	1.69	69.6	53.9	76.7	2.05
	S.D.....	.45	9.0	23.5	19.1	.57

means of two estimates given each student by two former instructors, rendered numerical by calling A, 1; B, 2; C, 3; and D, 4 (A representing the highest rating). This table is not complete in itself but must be supplemented by the statistical treatment to determine if the differences are significant, as shown in Table XXXI.

TABLE XXXI

RELIABILITIES OF DIFFERENCES[2] BETWEEN SECTIONS A AND B IN PHYSIOLOGY
(Pairs Only)

	Freshman Honor Points	Miller Test	Sigma X Test	Average Honor Points	Industry
Difference in means ÷ P.E. of difference....	.16‡	.26‡	2.5§	1.6§	.6§
Difference in S.D.'s ÷ P.E. of difference....	.2§	.1§	2.0§	.1§	.4‡

There are no significant differences.

According to this treatment the groups of pairs in Sections A and B were found comparable, as might be expected when one considers the method used in pairing them. Assuming that the ratings from psychological tests and the ratings in school subjects shown by honor points furnish measurements that may be treated as approximately accurate, we had two sections that might be considered equal in these respects, both as to all students in the sections and as to the groups of pairs.

What remained to be done was to show the ratings of the students in achievement in physiology. It was assumed that these

[2] Thorndike, *loc. cit.*

ratings were suitable to subject to the statistical treatment that was used. They were not scaled ratings but were much more complete and carefully handled than the average run of ratings in school subjects. The ratings used were as follows: a mid-quarter objective test consisting of 70 true-false and 55 uncompleted statement items, 125 in all; a final, written objective test of 250 similar objective items; a term laboratory rating on the laboratory book handed in by the student, judged by the two laboratory instructors cooperating; a final oral rating, which was the mean rating given by any two of the four instructors selected by the student; a final laboratory mark, which was the mean estimate of two tests: a "set-up" and a "tracing"; and the mean of 19 five-minute quizzes given during the quarter. The total, as given in the tables that follow, is a composite, made up as follows: each rating was reduced to a percentage system, representing points in a hundred; then each was multiplied by factors, the mid-quarter test by 1, the final written test by 4, the final oral one by 1, the average of the five-minute quizzes by 2, the term laboratory test by 1, and final laboratory test by 1, to make a possible total of 1,000 points. The total was, thus, a certain number of points out of a thousand possible points. "Total objective tests" in the tables indicates added raw scores of the mid-quarter and final tests. Table XXXII gives these ratings for all students in the sections. Table XXXIII shows the reliability of the differences.

TABLE XXXII

SUMMARY OF ACHIEVEMENT RATINGS OF STUDENTS, SECTIONS A AND B IN PHYSIOLOGY

(All Students)

		RATINGS							GRAND TOTAL
		Mid-Quarter %	Final Written %	Total Obj. Score	Final Oral %	5-Min. Quiz %	Term Lab. %	Final Lab. %	
Section A	Mean.....	71.3	59.2	236.7	74.2	65.1	77.3	38.1	607.8
	S.D.....	11.7	14.5	45.6	9.9	12.0	12.3	25.5	106.9
Section B	Mean.....	61.7	55.7	220.2	76.8	48.7	66.2	33.1	557.7
	S.D.....	12.5	13.5	48.3	10.4	13.9	14.7	21.0	108.5

TABLE XXXIII

RELIABILITIES OF DIFFERENCES BETWEEN SECTIONS A AND B IN PHYSIOLOGY
(All Students)

	Ratings							Grand Total
	Mid-Quarter %	Final Written %	Total Obj. Score	Final Oral %	5-Min. Quiz %	Term Lab. %	Final Lab. %	
Difference in means ÷ P.E. of difference	4.8‡	1.5‡	2.1‡	1.4§	7.5‡	4.4‡	1.2‡	2.7‡
Difference in S.D.'s ÷ P.E. of difference	.5§	.6‡	.5§	.4§	1.2§	1.5§	1.6‡	.1§

There appeared considerable differences in favor of Section A, which had seven and one-half hours of laboratory, in all ratings except the final oral.

There were significant differences in the means of the mid-quarter test, the means of the five-minute quizzes, and the term laboratory work. The differences in the means of the total objective scores and of the total were also large and might be considered significant by some authorities. There were no significant differences in the S.D.'s. The final oral showed the only difference that was in favor of Section B. This was the mean of two estimates of instructors given after each had engaged in a short conference with the student and was perhaps determined to a considerable extent by personality judgments. Considerable evidence is indicated in favor of seven and one-half hours of laboratory work a week, as contrasted with five hours of laboratory work plus two and one-half hours of library work, at least as the latter plan was conducted. It cannot, however, be conceded that the laboratory and the library work with Section B might not possibly have been varied so as to get better results. Scientific candor and accuracy make it necessary to say that the conclusions drawn apply only to the conditions as they existed. If exactly the same conditions were to be duplicated and the same measurements used, there is no reason to believe that the results would be different.

Curiously enough, when the data for the pairs alone were used the results were not as conclusive. Table XXXIV (a) gives the data on achievement for the pairs alone.

TABLE XXXIV (a)

COMPARISONS OF ACHIEVEMENT RATINGS, SECTIONS A AND B IN PHYSIOLOGY, MEANS AND STANDARD DEVIATIONS

(Pairs Only)

		RATINGS							GRAND TOTAL
		Mid-Quarter %	Final Written %	Total Obj. Score	Final Oral %	5-Min. Quiz %	Term Lab. %	Final Lab. %	
Section A	Mean.....	70.0	57.1	229.6	73.3	53.4	78.2	36.0	595.8
	S.D.....	12.9	13.4	44.3	9.6	11.3	9.8	24.2	95.4
Section B	Mean.....	63.8	55.1	217.3	76.1	49.8	66.2	32.2	557.2
	S.D.....	12.0	12.9	43.4	11.3	12.7	13.6	21.5	101.5

TABLE XXXIV (b)*

COMPARISONS OF ACHIEVEMENT RATINGS, SECTIONS A AND B IN PHYSIOLOGY

Number in Section B Exceeding Paired Companions in Section A

(Twenty Pairs)

	NUMBER
Mid-quarter test	6
Final written test	9
Total objective tests	7
Final oral test	12
Mean of weekly quizzes	7
Term laboratory mark	6
Final laboratory test	9
Final rating	7

* Individual students.

Considering the pairs as individuals, the data of comparison are given in Table XXXIV (b). These show a preponderance of evidence in favor of Section A in all criteria except the final oral. More than half of Section A had higher ratings in every other criterion than their paired companions in Section B. There were from six to twelve students in Section B, however, who did better than their paired companions in Section A.

Table XXXV gives the data on the reliability of the differences. There were similar differences noted in favor of Section A, but they were not so great. According to the same criterion for significance the only significant difference was in the term laboratory grade, which was the rating on the notebook.

TABLE XXXV

RELIABILITIES OF DIFFERENCES, SECTIONS A AND B, IN PHYSIOLOGY
(Pairs Only)

	Ratings							Grand Total
	Mid-Quarter %	Final Written %	Total Obj. Score	Final Oral %	5-Min. Quiz %	Term Lab. %	Final Lab. %	
Difference in means ÷ P.E. of difference	2.4‡	.7‡	1.3‡	1.3§	1.6‡	4.8‡	.8‡	1.8‡
Difference in S.D.'s ÷ P.E. of difference	.5‡	.2‡	.1‡	1.1§	.8§	2.2§	.8‡	.4§

It seems, then, that there were certain individuals who did get more from five hours of laboratory work and two and one-half hours of library work than they probably would have gained from seven and one-half hours of laboratory work, at least as measured by the ratings used. Is it possible in some way to pick them out by some preliminary testing and allow them some degree of freedom in their choice of the time given to laboratory work?

Table XXXVI gives certain coefficients of correlation.

TABLE XXXVI

COEFFICIENTS OF CORRELATION*

Achievement	Other Rating	Coefficient	P.E.	No. of Cases
Physiology (total objective score)	Achievement (total points)	.93	.01	61
Physiology (total objective score)	Freshman honor points	.68	.04	62
Physiology	Miller test	.38	.07	62
Physiology	Achievement anatomy	.50	.07	50
Physiology	Average honor points, required subjects	.49	.06	62
Average honor points, required subjects	Average honor points, all subjects	.79	.02	131

* Otis Correlation Chart used.

The first shows the close relationship between the objective test ratings and the final rating. The next four show rather low rela-

tionships between achievement in physiology and other ratings which might be considered as prognostic of achievement in physiology. Freshman honor points are most prognostic. It may be added that this is the highest correlation obtained in any of the experiments described in this book, used for determining prognosis value. The last shows a fairly high relationship between average honor points in all subjects and in the required subjects. The question arose whether or not it was best to use average honor points in all premedical subjects or only in the required subjects. The correlation was computed to find out whether there would be a high correlation one with the other. The coefficient obtained shows fairly high relationship. As it turned out, freshman honor points were more predictive.

CHAPTER V

EFFECT OF REDUCTION OF TIME GIVEN TO LABORATORY IN HUMAN PHYSIOLOGY

This is a companion experiment to that which was described in Chapter IV. The object in this experiment was to determine the effect of lessening the time given to laboratory work in which two and one-half hours per week of laboratory work were replaced with eleven hours of lecture work per quarter on bio-physics. While in the other experiment library work bearing on the subject of physiology was substituted, the time requirements being unchanged, in this experiment only eleven hours of lecture work on a distantly related subject were substituted for fifty hours of laboratory work. This was an attempt to eliminate some of the laboratory work, if it could be shown that the efficiency of instruction was not lowered. Or, to put it another way, it was an attempt to find the optimal amount of laboratory work desirable in achieving certain results that the instructional staff had in mind.

Sections C and D are the sections involved in this experiment. As with the sections involved in the experiment described in the preceding chapter, there were five lectures given each week of the quarter to both Sections C and D by Dr. Scott, the lectures being the same to both sections. There were also two quizzes per week under instructors Greisheimer and King, the two instructors reversing at mid-quarter so that they would have each section half of the time. Laboratory Division C, was in charge of Instructor Loucks the first half-quarter, while Division D was under Instructor King. During the second half-quarter, the instructors reversed so that each section would have each instructor half of the quarter. Division C, which may be called the control group, had seven and one-half hours a week of laboratory work, while Division D had but five hours a week of laboratory work, plus a total of eleven hours per quarter of lecture work on bio-physics under Dr. Stenstrom.

For this experiment the students in the sections were drawn at random. There was no attempt made to get paired students or equated sections at the beginning. Section D, for administrative

convenience, was composed of irregular students and was also somewhat larger than Section C. There were also some records missing for certain students in this section. After the sections had been organized, however, it was possible to pick out from Section D fourteen students who apparently paired quite well with fourteen students in Section C. This pairing was on the basis of past scholarship, represented by average honor points in all premedical work and the rating in the Miller mental ability test for graduate students. These were the only complete records available. After the work had started, two of the paired students in Section D dropped out, so that finally there were but twelve pairs to compare in the two sections.

Table XXXVII gives the data in regard to the twelve pairs of students, adding the scores in the Sigma X reading test for com-

TABLE XXXVII

SUMMARY OF DATA FOR THE EQUATION OF GROUPS, SECTIONS C AND D IN PHYSIOLOGY

(Pairs Only)

	AVERAGE HONOR POINTS	MILLER TEST	SIGMA X TEST
Section C* Mean	1.27	62.8	68.0
S.D.	.26	14.9	21.5
Section D† Mean	1.26	62.0	69.0
S.D.	.27	13.9	16.3

* Laboratory, 7½ hours.

† Laboratory, 5 hours; lecture on bio-physics, 11 hours per quarter.

TABLE XXXVIII

RELIABILITIES OF DIFFERENCES BETWEEN SECTIONS C AND D IN PHYSIOLOGY*

(Pairs Only)

	AVERAGE HONOR POINTS	MILLER TEST	SIGMA X TEST
Difference in means ÷ P.E. of difference	.1††	.2††	.2¶¶
Difference in S.D.'s ÷ P.E. of difference	.2¶¶	.4††	1.4††

There were no significant differences in the sections.

* In this and succeeding tables †† indicates C greater and ¶¶ indicates D greater.

parison. While 12 is probably not a large enough number to handle statistically, Table XXXVIII is given to show the treatment used in determining the reliability of the differences in the groups of paired students.

TABLE XXXIX

COMPARISONS TO SHOW HOW NEARLY PAIRS IN SECTION C ARE RANDOM SAMPLINGS OF THAT SECTION

	Av. H. P.	Miller Test	Sigma X Test
Difference in means ÷ P.E. of difference	.5	.5	.3
Difference in S.D.'s ÷ P.E. of difference	1.4	.2	.0

Twelve pairs are compared with all students as given in Table XLI.

TABLE XL

COMPARISONS TO SHOW HOW NEARLY PAIRS IN SECTION D ARE RANDOM SAMPLINGS OF THAT SECTION

	Av. H. P.	Miller Test	Sigma X Test
Difference in means ÷ P.E. of difference	2.4	.8	.7
Difference in S.D.'s ÷ P.E. of difference	3.6	.6	.7

Twelve pairs are compared with all students as given in Table XLI.

A further procedure is shown in Tables XXXIX and XL, having for its purpose the determination of whether or not the pairs selected from Sections C and D are representative of the sections; i.e., whether they represented random samples of the sections, or whether they represented selective groups within the sections. It was desirable that the data in regard to the pairs be representative of the sections as a whole in order to draw conclusions from the pairs which would probably apply as well to the sections considered as wholes. The data in Table XXXIX show that the twelve students taken from Section C as members of the pairs were good samples of the group as a whole because there are no significant differences between the means or standard deviations of the twelve and of the section as a whole. The data in Table XL show that the

twelve students from Section D represent a fair sample of the whole section. There is a significant difference between the standard deviations of the group of twelve and the whole section in average honor points but not in the mental ability or the reading tests. The group of twelve has a somewhat lower mean scholarship than the whole section.

It must be noted, however, that not quite all members of the sections are represented in the data because of incomplete records. The data for those having complete records are shown in Table XLI. Forty-four students completed the course in Section D and

TABLE XLI

Summary of Data for the Comparison of Groups, Sections C and D in Physiology

		Students
Section C		
Miller test, mean	64.4	28
Miller test, S.D.	15.4	28
Total average honor points, mean	1.30	27
Total average honor points, S.D.	.33	27
Sigma X test, mean	69.6	27
Sigma X test, S.D.	21.6	27
Section D		
Miller test, mean	64.57	35
Miller test, S.D.	12.6	35
Total average honor points, mean	1.43	39
Total average honor points, S.D.	.45	39
Sigma X test, mean	71.7	35
Sigma X test, S.D.	18.2	35

Some records are missing.

twenty-nine in Section C. The number whose records were incomplete may, therefore, be noted by examining Table XLI. The data in Tables XXXIX and XL are based on the number given in Table XLI.

The achievement of the groups of pairs in the two sections was estimated exactly as the achievement of Sections A and B described in Chapter IV. Tables XLII (a) and XLII (b) give the data for the pairs from the two sections. Section D was somewhat

TABLE XLII (a)

COMPARISONS OF ACHIEVEMENT RATINGS, SECTIONS C AND D IN PHYSIOLOGY, MEANS AND STANDARD DEVIATIONS
(Pairs Only)

		Ratings							Grand Total
		Mid-Quarter %	Final Written %	Total Obj. Score	Final Oral %	5-Min. Quiz %	Term Lab. %	Final Lab. %	
Section C	Mean.....	66.3	48.7	204.4	71.2	48.2	69.1	45.7	545.17
	S.D.....	15.7	13.2	51.5	12.3	7.8	10.3	20.6	99.2
Section D	Mean.....	68.6	51.4	207.3	75.1	55.9	71.2	35.2	567.7
	S.D.....	9.5	10.4	30.5	8.4	7.9	6.5	25.8	69.5

TABLE XLII (b)*

COMPARISONS OF ACHIEVEMENT RATINGS, SECTIONS C AND D IN PHYSIOLOGY
Number in Section C Exceeding Paired Companions in Section D
(Twelve Pairs)

	Number
Mid-quarter test	5
Final written test	6
Final objective test	5
Final oral test	6
Mean of weekly quizzes	4
Term laboratory mark	6
Final laboratory test	9
Final rating	5

* Individual students.

higher in all criteria except the final laboratory rating. A greater number of students in Section D exceeded their paired companions in Section C in all criteria except the final laboratory, final written, final oral, and the term laboratory ratings. Only in the final laboratory did a greater number of students in Section C exceed their paired companions in Section D.

Table XLII (c) shows the statistical treatment to ascertain the significance of the differences. According to this treatment there was a significant difference only in the means of the nineteen five-minute quizzes. The variabilities of the group from Section C, as indicated by the standard deviations, are greater

TABLE XLII (c)

RELIABILITIES OF THE DIFFERENCES BETWEEN SECTIONS C AND D IN PHYSIOLOGY
(Pairs Only)

	Ratings							
	Mid-Quarter Per Cent	Final Written Per Cent	Total Objective Score	Final Oral Per Cent	5-Minute Quiz Per Cent	Term Laboratory Per Cent	Final Laboratory Per Cent	Grand Total
Difference in means....	2.3±3.6¶¶	2.7±3.3¶¶	2.9±11.5¶¶	4.9±2.9¶¶	7.7±2.1¶¶	2.1±2.4¶¶	10.5±6.4††	22.5±23.5¶¶
Difference ÷ P.E. of difference.....	.6	.8	.2	1.7	3.7	.9	1.6	1.0
Difference in S.D.'s.....	6.2±2.6††	2.8±2.3††	21.0±8.2††	3.9±2.1††	.1±1.6¶¶	3.8±1.8††	5.2±4.6¶¶	29.7±10.2††
Difference ÷ P.E. of difference.....	2.4	1.2	2.6	1.9	.1	2.1	1.1	2.9

than for Section D. The conclusions drawn from the data certainly did not show any evidence in favor of Section C. What little evidence there was pointed toward higher achievement by Section D. It would have been highly desirable, however, to have had a greater number of students than twelve on which to base conclusions. A further study of the individuals in the two groups might have been worth while and disclosed factors that should be considered in drawing conclusions. There was the factor of difference in the sizes of the sections, which was not eliminated in the data. Since the conclusions are exactly opposite to those given for Sections A and B, described in Chapter III, conclusions must be but tentative pending further experimentation. The conclusions drawn from the experiment with Sections A and B were based on more complete data and were probably more valid and reliable than any conclusions drawn from the data on Sections C and D. It did not seem wise, however, to discard these data, although the results should receive further study.

Subsequent Data

The records for the pairs of Sections C and D have been examined more in detail since the above experiment was performed and reveal some possible inaccuracies in drawing conclusions from such small numbers of paired students. The data show more irregularities in Section D than in Section C. For example, but one student in Section C had had the course in physiology previously, while four in Section D were repeating it. One student in Section C had never taken anatomy. All students in Section D had taken a course in Physiology 136, an unlisted course, which may have had some effect on their achievement in Physiology 103, the one under consideration here. Moreover, two in Section D had taken the course in Physiology 104, which ordinarily follows Physiology 103.

These factors all bear directly on achievement in Physiology 103 and should receive attention in any evaluation of teaching procedures. They all have a tendency to increase measurable achievement for Section D. Therefore, no valid conclusions as to the efficacy of the procedures with the sections, as were primarily under consideration, can be drawn. It is clear that there are here other variables than those of amount of time in laboratory work.

It is particularly worth while, however, to point out that especially with such small groups it may be unwise to try to draw final conclusions on the basis of pairing students. The factors determining achievement are so varied that no two students can be paired in all of them. With fairly large-sized groups usually no marked, unknown variation of a single individual in an especially valid factor would be great enough to affect, noticeably, the group average. In such a case there would have to be, necessarily, a great difference in many individuals to make the variation apparent. Evidence here points, then, to the value of having fairly large groups and making comparisons on a group basis. Individuals may then receive separate study.

PART III

STUDIES IN THE PHYSICS DEPARTMENT

CHAPTER VI

PRELIMINARY TESTING IN MECHANICS

This section deals with several phases of an attempt in the Department of Physics to determine, experimentally, differences in the achievement of certain architectural students and of other students enrolled in the course in mechanics. All students in this course were required to attend three lectures, one quiz, and one double-period laboratory class each week, except those students who were majoring in architecture. The architectural students were excused from the laboratory work. What is the effect, as regards achievement, of their lack of laboratory work? Or, in other words, what do the students who engage in the laboratory work gain that is not gained by the students in architecture?

In this experiment an attempt was made to get bases for pairing students in addition to and differing from those used in the anatomy and physiology experiments. In the latter no initial measurements were made in anatomy or physiology, respectively. The paired students were compared in previous school ratings and in the results of psychological tests, but they were not compared initially in their preliminary knowledge of anatomy or physiology, respectively. In the experiment in physics two quite comprehensive preliminary tests covering the field of high school physics were given. These consisted of some high school tests devised by the writer and used for several years in measuring achievement in high school physics. Norms based on the work of from 337 to 541 students in Minneapolis high schools, which were available for these tests, were used in making comparisons. The first test, consisting of Hurd Tests II, III, and Part I of Test VI, was given to the three large sections, including all students, on the first day of the quarter; the second, consisting of Hurd Tests VIII, X, Part II of Test VI, and Part II of Test IX, was given during the laboratory periods of the first week. Tables XLIII (a) and XLIII (b) show the distribution of scores in the two tests. The correlation for 288 students between the first and second tests was found to be $.77 \pm .016$ (Otis Correlation Chart). This showed considerable degree of relationship between the tests.

TABLE XLIII (a)

DISTRIBUTION SCORES IN HURD PHYSICS TEST*
First Test Composite of Hurd Tests II, III, and Part I of Test VI
(All Students)

SCORE	FREQUENCY	SCORE	FREQUENCY
0– 4.9	9	30–34.9	42
5– 9.9	27	35–39.9	28
10–14.9	43	40–44.9	13
15–19.9	62	45–49.9	2
20–24.9	88	50–54.9	3
25–29.9	62		
			379

Possible Score—62. Median, 22.8. Q, 6.6.

* Given to students enrolled in Mechanics, fall quarter.

TABLE XLIII (b)

DISTRIBUTION SCORES IN HURD PHYSICS TEST*
Second Test Composite of Hurd Tests VIII, X, Part II of Test VI, and Part II of Test IX
(All Students)

SCORE	FREQUENCY	SCORE	FREQUENCY
0– 4.9	2	55– 59.9	25
5– 9.9	11	60– 64.9	26
10–14.9	7	65– 69.9	24
15–19.9	13	70– 74.9	22
20–24.9	16	75– 79.9	15
25–29.9	27	80– 84.9	10
30–34.9	21	85– 89.9	3
35–39.9	23	90– 94.9	1
40–44.9	26	95– 99.9	2
45–49.9	31	100–104.9	1
50–54.9	29	105–109.9	1
			336

Possible Score—129. Median, 48.5. Q, 16.1.

* Given to students enrolled in Mechanics, fall quarter.

The administration of these tests made possible some comparisons which are directly related to the study of architectural students but which also are of themselves interesting and worth while.

TABLE XLIII (c)

COMPARISONS OF PRELIMINARY TEST SCORES IN MECHANICS OF STUDENTS IN DIFFERENT COLLEGE DEPARTMENTS

(First and Second Tests Combined)

	ENGINEERING	CHEMISTRY	PREMEDICAL AND DENTISTRY	EDUCATION
Mean score	79.84	75.06	53.71	55.0
Standard deviation	26.4	25.0	27.2	31.7
Number having had high school physics	31	8	18	5
Per cent having had high school physics	20	17	29	50
Number in group	152	46	62	10

Possible Score—191.

For example, the question of the ratings of various groups of students enrolled in mechanics in the pre-test on subject matter covered in high school physics arose. Table XLIII (c) presents some summary data comparing students in engineering, chemistry, medicine, and dentistry, and education in the 191 preliminary test items.

Students in engineering rated higher in preliminary test scores, on the average, than the other groups for which data are given. It would have been desirable to separate the engineers in Table XLIII (c); at least to segregate the architects. None of the latter, however, took the second part of the pre-test, due to their absence from laboratory sections where this test was administered. They were compared afterward in the first pre-test. These comparisons are shown in Chapter IX.

The data given in Table XLIII (c) immediately brought up the question of mental ability ratings. While no mental tests were given to the classes in mechanics, there were records for many of these students of scores given at the beginning of the freshman year to all entering students, who were given a comprehensive test of the type usually called intelligence tests. Table XLIV presents a summary of these records for the same groups as are included in Table XLIII (c), with the exception of those in education. A large number of students in education do not take these tests.

TABLE XLIV

SUMMARY OF PERCENTILE SCORES IN FRESHMAN MENTAL TEST

	ENGINEERING	CHEMISTRY	PREMED. & DENT.
Mean score	56.0	69.3	55.1
Standard deviation	27.1	25.7	26.9
Number of students	160	35	67

The chemists were somewhat superior, apparently, with little difference between the engineers and the predentals and premedics. All three groups showed some selection on the basis of mental ratings, as the means are all higher than the 50-percentile (the middle of the total group of all students taking the test). This introduced the question of relationships between mental ratings and ratings on the Hurd tests. Did these pre-tests measure general mental ability? The following correlations were computed (using the Otis Charts) for only those who had studied physics in high school. One hundred students were used for each correlation.

1. Raw scores, first test and freshman percentile intelligence ..... $.14 \pm .065$
2. Raw scores, second test and freshman percentile intelligence ..... $.20 \pm .06$
3. Raw scores, both tests (added scores) and freshman percentile intelligence..... $.39 \pm .06$

The correlations are all low, showing that the preliminary tests cannot be considered as tests of mental ability. Their value as prognostic tests for mechanics is discussed in a later chapter.

THE VALUE OF SCALING INDIVIDUAL TEST QUESTIONS

If the preliminary test scores were to be used as bases of pairing students in architecture with those of other groups taking laboratory work, a very significant question was the manner in which these test scores should be expressed. Should raw scores be used; should the scaled value of each question be computed to be used in finding final score values; or should final raw score values be expressed in some standard values such as units of probable error or standard deviation? Some study of these alternatives was possible because of records available from the use of these tests with pupils

of Minneapolis high schools. The scale values of the items in the Hurd Tests II, III, VI, VIII, and X, using the Rugg technique of finding a scale value for each test item, were computed on the basis of the administration of these tests to several hundred of these high school pupils in the Minneapolis schools during the last three or four years. The difficulty of each test item was determined by finding the number of pupils failing on each item, computing the per cent that each of these numbers was of the total number of pupils taking the test, and getting the value of the item in standard deviation percentiles from the Rugg table.[1]

The first hundred papers from each set of the tests, as given to the students in mechanics, were then rescored, using these weighted values, and the values thus obtained added for a final score. This was done for Tests II, III, VIII, and X; the scores thus obtained were correlated with the raw scores used in the preceding tables. The following correlations were obtained:

1. Test II.....$r = .98 \pm .003$
2. Test III.....$r = .97 \pm .005$
3. Test III (revised raw scores, 1 for each whole item) $r = .99 \pm .001$
4. Test VIII.....$r = .98 \pm .003$
5. Test X.....$r = .98 \pm .003$

Test III is regularly scored by giving a credit of one point for each blank filled out correctly. The revised score was on the basis of one point for each whole item. The Otis Correlation Chart was used for determining these correlations. These correlations are so high that they give authority for using the raw scores for such work as has been indicated in the preceding tables. Table XLV shows the technique used in getting the scaled values for the test items in Test X. The other tests were treated in the same manner.

The Scaling of the Preliminary Tests

The use of the Rugg technique was laborious and made the tests no more valid in ranking students. There were, also, other reasons which made its use unessential. It could not be proved that the scale value of any given item should be determined by the number of persons failing to give a correct response to it. This failure might not be due to inherent difficulty but to the infrequency with which

[1] Rugg, H. O., *Statistical Methods Applied to Education*, p. 392.

TABLE XLV

COMPUTATION OF S.D. PERCENTILES, TEST X, ELECTRICITY AND MAGNETISM
(380 Pupils)

	TEST ITEM	NO. CORRECT ANSWERS	PER CENT FAILED	S.D. PERCENTILE[1]
Part I	1	375	1.3	8.6
	2	371	2.4	12.4
	3	363	4.5	17.4
	4	329	13.4	28.4
	5	335	11.8	26.9
	6	332	12.6	27.7
	7	331	12.9	27.8
	8	311	18.2	32.3
	9	286	24.7	36.7
	10	238	37.4	43.9
	11	263	30.8	40.8
	12	250	34.2	42.2
	13	248	34.7	42.4
	14	226	40.5	45.5
	15	206	45.8	48.2
	16	163	57.1	53.2
	17	151	60.3	54.9
	18	156	58.9	54.2
	19	164	56.8	53.1
	20	109	71.3	60.9
	21	96	74.8	63.0
	22	166	56.3	52.9
	23	131	65.5	57.7
	24	95	75.0	63.1
	25	107	71.8	61.2
	26	137	63.9	56.8
	27	119	68.7	59.4
	28	84	77.9	65.0
	29	46	87.9	72.8
	30	71	81.3	67.3
	31	49	87.1	72.0
	32	61	83.9	69.4
Part II	1	316	16.8	31.3
	2	173	54.5	51.9
	3	155	59.2	54.3
	4	149	66.1	58.0
	5	159	58.2	53.8
	6	128	66.3	58.1
	7	116	69.5	59.8
	8	97	74.5	62.8
	9	104	72.6	61.7
	10	69	79.2	65.9

Possible Score—42.

[1] Rugg, *loc. cit.*

it entered the experience of people in general. In other words, it might be an unimportant element, obtaining a high scale value merely on account of people never having considered it or having been exposed to it. Or it might have varying difficulty with various individuals, thus making every such item of different difficulty with different individuals. Without entering into any further discussion of the value of this technique in scaling, it is enough to say that there appeared no pressing reason why it should be used.

There was, however, the impression that a method of expressing scores in some standard fashion would be desirable. For example, if every test or other rating were expressed in a common manner, it would be more easy of interpretation; i.e., a person's position in a group could be more easily seen at a glance, and a raw score might be given a more true value. It was apparent that a difference in raw score of 1 near the middle of a distribution was not as significant as the same amount near the end of a distribution. If the concept of the normality of distribution of people in measures of this kind were true, as seemed evident from much factual data so far obtained, it seemed wise to express raw scores in some of the functions of a normal distribution. The theory was that the nearer a person comes to making a perfect score, the more difficult it becomes; therefore a difference of 1 near the high end is a more significant achievement than a difference of 1 at the middle. A similar logic was applicable for the low end of the distribution.

Hurd Tests II, III, VI, VIII, and X were finally treated by means of the McCall[3] T technique, the value of a given raw score being expressed in terms of a normal distribution in units of standard deviation. The percentage which the number exceeding plus one-half reaching a given raw score was of the total number of cases served as a basis for obtaining scale values. These values were obtained from a table expressing such percentages in fractions of standard deviations (the rating of 50 is at the middle of the scale). The tests were originally given to about 100 pupils in North High School, Minneapolis. The items were later arranged in order of difficulty on the basis of the 100 taking the test. The item getting the largest number of correct responses was put first; the one getting the fewest number, last. Later, the test was given

[3] W. A. McCall, *How to Measure in Education*, chap. ix.

TABLE XLVI

First Test Given to Entering Classes in Physics—Composite of Hurd Tests II, III and Part I of Test VI

Scores	Tallied Distribution	Frequency	No. Exceed. +½ Reach.	Per Cent Exceed. +½ Reach.	Scale[1] Value
0– 2.9	IIII	4	377	99.47	24.5
3– 5.9	IIIIIIIIIII	12	369	97.36	30.6
6– 8.9	IIIIIIIIIIIIII	15	355.5	93.80	34.6
9–11.9	IIIIIIIIIIIIIIIII	18	339	89.45	37.5
12–14.9	IIIIIIIIIIIIIIIIIIIIIIIIIIII	30	315	83.11	40.4
15–17.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	36	282	74.41	43.4
18–20.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	47	240.5	63.46	46.5
21–23.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	49	192.5	50.79	49.3
24–26.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	43	146.5	38.65	52.8
27–29.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	37	106.5	28.10	55.8
30–32.9	IIIIIIIIIIIIIIIIIII	22	77.0	20.32	58.2
33–35.9	IIIIIIIIIIIIIIIIIIIII	24	54.0	14.25	60.7
36–38.9	IIIIIIIIIIIIIII	17	33.5	8.84	63.5
39–41.9	IIIIIIIIIIII	14	18.0	4.75	66.7
42–44.9	IIIIII	6	8.0	2.11	70.3
45–47.9		0	5.0	1.32	72.2
48–50.9	III	3	3.5	.92	73.5
51–53.9	I	1	1.5	.40	76.5
54–56.9	I	1	.5	.13	80.0
		379			

Possible Score—62.

[1] McCall, *loc. cit.*

to other students in Minneapolis high schools, the rank order of the questions remaining very nearly as at first. The total score values might well be called "M" scores, for they employ the McCall T technique. The middle of the scale, 50, however, does not represent the median ability of a twelve-year-old as a correct T score should, but, rather, merely the median ability of the group of high school pupils who took the test. The scaled values of Test X, for example, range from 20 to 80.5, the range of values for the other tests being about the same. This represents a range of about three standard deviations above and below the mean, respectively, containing approximately 99.7 per cent of an ideal group.

These M scores were used in some further comparisons. Table XLVI shows a distribution of scores and the method of obtaining M values for the first composite preliminary test given to students in mechanics. Its use as a prognosis test for mechanics is discussed in Chapter VIII.

Validity and Reliability of Preliminary Tests

If the preliminary tests were to be used as bases of prognosis for pairing students in equated groups, it was necessary to establish their validity and reliability; at least it was desirable to know to what extent they were valid and reliable. Some data were available on these questions. Probably as good a measure of validity as can be obtained is that shown by the differences in ratings on the tests by those students who have studied high school physics, and those who have not. These differences were so marked that the argument for the validity of the tests seems quite convincing. More evidence has been given in studies heretofore published showing decided gains in the scores obtained by students on some of these tests before studying and after studying the subject.[5] In previous studies by the writer[6] a reliability coefficient of $.69 \pm .04$ was obtained between scores on Part II, Test IX and its repetition

[5] A. W. Hurd, A Tropical Method Compared with a Problem Method in the Subject of Heat in Physics (Master's thesis, University of Minnesota, 1924); and "Subject Matter versus Character Traits," *Journal of Educational Method*, 5: 147-154 (1925).

[6] *Ibid.*

fourteen weeks later. Probably the correlation would be much higher if the elapsed time had not been so great. From the results of Tests II, III, and X, as given to the students in mechanics, scores were obtained on the even-numbered and odd-numbered questions, respectively, and correlated for each test. The correlations are as follows: Test II—.51±.05; Test III—.70±.03; Test X—.86±.017. These correlations may be corrected for a more accurate coefficient by the use of the Spearman-Brown formula,[7]

$$r_x = \frac{2r_h}{1 + r_h}.$$

The coefficients revised are as follows: Test II—.675±.04; Test III—.82±.02; Test X—.92±.01. The small number of questions answered by the university students probably accounts for a lack of higher coefficients. Further evidence, though indirect, is deduced from the constancy in the means and standard deviations of the test scores obtained from classes for several years.

There is room for further work in reliability investigation; but the repetition of a test merely for a reliability index is not inspiring to students and in regular teaching activities is somewhat difficult to get. What evidence there is points to high reliability when it is considered that the reliability coefficients obtained are based on so few test items.[8]

It is pertinent to add that the tests cover informational and problem material. There are many other phases of instruction which have not as yet been measured, upon which it is very much worth while to reflect. The writer has done some work on tests to measure ability to pick out significant data, ability to deduce applications from principles, and ability to generalize.[9] School instruction in the future may be planned to develop such abilities directly. The present tests are more conventional.

The Achievement of High School Students in Physics

Were there preliminary differences between those students who had studied high school physics and those who had not? If there

[7] Henry E. Garrett, *Statistics in Psychology and Education,* p. 271.

[8] Since this was written, the reliability correlation has been computed from the administration of eighty-seven items selected from these same preliminary tests. The coefficient obtained was .94±.02.

[9] Unpublished data.

were, they should receive consideration in pairing students for the study of the relative achievements of architectural students and other contrasted groups. Data applicable to this question were gathered by tabulating test scores for the two groups. As tentative test norms had been established by the administration of these tests, covering the field of high school physics to several hundred high school pupils, the scores were tabulated for each test for comparison with these norms. Total scores in all the pre-tests were also tabulated. Foley[10] published some data in 1922 showing great lack of ability of students who had taken high school physics, in solving some problems in the field. His conclusions indicate the failure of the high school course in teaching high school physics. The data presented in Table XLVII (a) and Table XLVII (b) show a great difference between the high school physics and non-physics student. Foley's data were based on a few problems, while these data are based on 191 items, some of them simple problems, others informational content.

This table shows the great superiority in every test of the students having had high school physics over those not having had it. Statistical comparisons, based on the means and standard deviations

TABLE XLVII (a)

DISTRIBUTION OF SCORES IN HURD PHYSICS TESTS* II, III, VI, VIII, X, AND A COMBINATION OF SEVERAL TESTS

	STUDENTS HAVING HAD H. S. PHYSICS					
	Test II	Test III	Test VI	Test VIII	Test X	Comb. Tests
Median	6.7	13.1	8.7	5.1	16.6	79.8
Median (norms)	10.8	17.4	12.8	12.9	20.1	
Score range	0–16	0–30	1–20	0–16	0–36	20–155
S. interquartile range	2.1	3.7	2.6	2.2	4.3	18.2
S. interquartile range (norm)	2.5	4.0	2.1	2.9	4.4	
Possible score	18	30	21	28	42	191
Number of high school pupils for norm	502	448	264	358	330	
Number of students in mechanics	297	301	237	272	266	229

*Given to students enrolled in mechanics, fall quarter, 1927.

[10] Arthur L. Foley, "The College Students' Knowledge of High School Physics," *School Science and Mathematics*, 22: 601-612.

TABLE XLVII (b)

DISTRIBUTION OF SCORES IN HURD PHYSICS TESTS* II, III, VI, VIII, X, AND A COMBINATION OF SEVERAL TESTS

	Students Not Having Had H. S. Physics					
	Test II	Test III	Test VI	Test VIII	Test X	Comb. Tests
Median	3.5	6.5	4.2	2.3	7.0	41.0
Median (norms)	10.8	17.4	12.8	12.9	20.1	
Score range	0–10	0–14	0–10	0–7	0–24	0–90
S. interquartile range	1.7	3.0	2.6	1.0	3.2	15.4
S. interquartile range (norm)	2.5	4.0	2.1	2.9	4.4	
Possible score	18	30	21	28	42	191
Number of high school pupils for norm	502	448	264	358	330	
Number of students in mechanics	84	83	60	66	70	62

* Given to students enrolled in mechanics, fall quarter, 1927.

TABLE XLVIII

COMPARISONS OF STUDENTS WHO HAD HIGH SCHOOL PHYSICS WITH THOSE WHO DID NOT

	H.S. Physics	Non-H.S. Physics
Mean score in all tests	80.65±1.14	42.50±1.79
Standard deviation	25.5	21.0
Coefficient of variability	.3	.5
Difference in means	38.15±2.12	
Difference ÷ P.E. of difference	(38.15÷2.12) = 18.0	

of the scores in the combined tests, are made in Table XLVIII. This statistical evidence shows the immense superiority of those having had high school physics.

Such evidence as this always introduces the question of mental ability ratings in their relation to achievement. Was the group having high school physics superior mentally? Data on this question were gathered from the freshman test percentiles referred to previously in Table XLVI. The tabulated distributions are shown in Table XLIX. There is shown no significant difference between the means of the two groups, though a tendency favoring the group having high school physics is noticeable.

TABLE XLIX

SUMMARY OF THE DISTRIBUTION OF PERCENTILE SCORES IN THE FRESHMAN MENTAL TEST*

STUDENTS HAVING HAD HIGH SCHOOL PHYSICS	STUDENTS NOT HAVING HAD HIGH SCHOOL PHYSICS
Mean..... 58.03 ± 1.9**	Mean..... 54.2 ± 3.6**
S. D..... 27.2	S. D..... 29.0

58.03 — 54.2 = 3.83 ± 4.1

3.83 ÷ 4.1 = .9‡‡

* In this and succeeding tables, ** indicates standard error of mean; and ‡‡ standard error of difference.

TABLE L

SUMMARY OF THE DISTRIBUTION OF COMBINED SCORES IN HURD PHYSICS TESTS*

STUDENTS FROM TWIN CITY PUBLIC HIGH SCHOOLS	STUDENTS FROM OTHER HIGH SCHOOLS
Mean..... 83.6 ± 2.7**	Mean..... 78.0 ± 2.2**
S. D..... 24.4	S. D..... 26.4

83.6 — 78.0 = 5.6

5.6 ÷ 3.5 = 1.6‡‡

Possible Score—191.

* Given to students enrolled in mechanics, fall quarter, 1927.

TABLE LI

SUMMARY OF THE DISTRIBUTION OF PERCENTILE SCORES IN THE FRESHMAN MENTAL TEST*

STUDENTS FROM TWIN CITY PUBLIC HIGH SCHOOLS	STUDENTS FROM HIGH SCHOOLS OUTSIDE TWIN CITIES
Mean..... 57.54 ± 2.62**	Mean..... 58.61 ± 2.37**
S. D..... 28.5	S. D..... 27.0

58.61 — 57.54 = 1.07 ± 3.53

1.07 ÷ 3.53 = .3‡‡

* Given to students enrolled in mechanics, fall quarter, 1927.

Did the students from high schools outside of the Twin Cities show different achievement standards than those within the Cities? Table L makes this comparison, while Table LI makes the similar comparison on freshman percentile mental ratings. The Twin City

high schools show a tendency toward higher achievement in general, though the difference is not statistically significant. There is much less difference in mental ability percentiles. The data show the test norms to be fairly representative of Minnesota high schools in general.

CHAPTER VII

MEASURES OF ACHIEVEMENT IN MECHANICS

Achievement in mechanics was estimated by averaging weekly conventional quizzes, three special objective tests, and the conventional final examination. One of the objective tests was given during the first half of the final examination period. Before this quarter, new-type objective examinations had not been used in the Physics Department. For this experiment, however, several of these new-type examinations were tried out.

The next to the last hour quiz period was given over to a new-type examination devised by Dr. J. W. Buchta of the Physics Department. It consisted of eighteen multiple choice questions, thirteen matching questions, and twenty-six uncompleted statement

TABLE LII

Distribution of Marks (Numerical) in Mechanics and M Values*

Numerical Mark	F	Scale Value
15–19	1	20.5
20–24	1	24.0
25–29	3	27.2
30–34	2	29.3
35–39	18	33.7
40–44	12	37.0
45–49	18	39.5
50–54	23	42.0
55–59	38	45.1
60–64	61	49.2
65–69	50	53.8
70–74	36	58.0
75–79	27	62.0
80–84	11	66.2
85–89	6	69.5
90–94	3	72.6
95–99	2	77.3
	312	

* McCall T Score technique used.

TABLE LIII

THE BUCHTA TEST SCALED ACCORDING TO THE MCCALL T SCORE TECHNIQUE

SCORE	TALLIED DISTRIBUTION	F	SCALE VALUE
15–17.9	IIIII	5	26.0
18–20.9	II	2	29.2
21–23.9	IIIIIIIIIIIIIII	15	33.1
24–26.9	IIIIIIIIIIIIIIIIIIII	21	37.4
27–29.9	IIIIIIIIIIIIIIIIIIIIIIIIIIII	28	40.8
30–32.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	47	44.5
33–35.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	56	49.0
36–38.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	53	53.5
39–41.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	36	58.8
42–44.9	IIIIIIIIIIIIIIIIIIII	23	61.4
45–47.9	IIIIIIIIIII	11	64.4
48–50.9	IIIIIIII	8	67.0
51–53.9	IIIIIIII	8	70.6
54–56.9	II	2	77.4
		315	

questions. The content covered the quarter's work. It was given in two forms that contained the same questions but in different order. These will hereafter be called Sets A and B. The last-hour quiz period was given over to a laboratory test of twenty-four uncompleted statement questions, supplemented by sixteen uncompleted statement text questions, and ten multiple choice questions, most of the last being statements of formulas. This examination was also administered in Sets A and B, which contained the same items but in different order. These tests were devised by Dr. H. A. Erikson, chairman of the department, Dr. Buchta, and the writer. The objective test occupying one-half of the final two-hour examination period was composed of twelve short answer questions and thirty-five uncompleted type questions devised by the same three persons. These three objective examinations may be considered together as a three-hour examination of 154 items.

It is to be noted that the students majoring in architecture were excused from the two-hour laboratory period a week, so that the laboratory test from their standpoint was unfair. The laboratory questions, however, were devised so that it would be possible for a student to get the necessary information from the text, although

there may have been advantages gained by other students from the laboratory work.

Table LII shows the distribution of the final numerical ratings given to 312 students enrolled in mechanics, together with the scale values. These were the students on the original list of all those taking the preliminary tests in physics who received final numerical ratings. Among these are included some who were repeating the lecture course for a second or third time but who may have received a passing mark in the laboratory work (these marks are recorded separately in the Registrar's office). The form of the distribution resembles a normal one. Of the number of students whose names

TABLE LIV

SECOND OBJECTIVE TEST SCALED WITH DISTRIBUTION OF RAW SCORES AND M VALUES

SCORE	TALLIED DISTRIBUTION	F	SCALE VALUE
6– 7.9	I	1	20.5
8– 9.9	I	1	24.2
10–11.9	I	1	26.0
12–13.9	I	1	27.2
14–15.9	IIIII	5	30.3
16–17.9	IIIIIIIIIII	11	33.2
18–19.9	IIIIIIIIIII	11	36.1
20–21.9	IIIIIIIIIIIIIIII	16	38.5
22–23.9	IIIIIIIIIIIIIIIII	17	40.9
24–25.9	IIIIIIIIIIIIIIIIIIIIIIII	24	43.1
26–27.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	32	45.7
28–29.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	31	48.5
30–31.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	34	51.2
32–33.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIII	30	53.9
34–35.9	IIIIIIIIIIIIIIIIIIIIIIIII	27	57.6
36–37.9	IIIIIIIIIIIIIIIIIIIII	23	59.4
38–39.9	IIIIIIIIIIIIIIII	17	62.3
40–41.9	IIIIIIIIIIII	13	65.5
42–43.9	IIIIIII	7	69.2
44–45.9	III	3	72.8
46	I	1	75.9
48	I	1	79.5
		307	

Irregulars omitted.

were on the original list—those taking either of the two preliminary tests—sixty-six cancelled the course during the quarter and 141 received either E (condition) or F (fail).

Table LIII shows the distribution for the Buchta test given during the next to the last quiz period, together with the scaled M values.

Table LIV shows the distribution of the scores in the laboratory and text quiz (second objective test) given during the last period. The students who had received a passing mark in laboratory at a previous time were omitted from the distribution.

Table LV shows the distribution of the scores in the final objective test as scored originally, giving one credit to each whole item together with the M scaled values.

TABLE LV

DISTRIBUTION OF THE FINAL OBJECTIVE TEST, AS SCORED ORIGINALLY IN MECHANICS, AND THE SCALED M SCORE VALUES

SCORE	TALLIED DISTRIBUTION	F	SCALE VALUE
0– 4.9			
5– 9.9	I	1	20.5
10– 14.9	II	2	25.1
15– 19.9	III	3	28.1
20– 24.9	IIIIIII	7	31.2
25– 29.9	IIII	4	33.4
30– 34.9	IIIIIIIIII	10	35.3
35– 39.9	IIIIIIIIIII	11	37.5
40– 44.9	IIIIIIIIIIIIIIIIIIII	21	39.9
45– 49.9	IIIIIIIIIIIIIIIIIIIIIIIII	27	42.7
50– 54.9	IIIIIIIIIIIIIIIIIIIIIII	25	45.2
55– 59.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	37	47.8
60– 64.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	42	51.0
65– 69.9	IIIIIIIIIIIIIIIIIIIIIII	26	53.9
70– 74.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	33	56.6
75– 79.9	IIIIIIIIIIIII	15	59.3
80– 84.9	IIIIIIIIIIIIIII	18	61.5
85– 89.9	IIIIIIIIIII	13	64.4
90– 94.9	IIIIIIIII	10	67.7
95– 99.9	IIII	4	71.5
100–104.9	III	3	76.0
		312	

Do Sets A and B Give Equivalent Results?

Table LVI shows the distributions in Sets A and B of the Buchta test. The means and standard deviations show that the two sets give similar results. Evidently, the rearrangement of the questions produces no serious disturbance in the distribution of scores. It will be recalled that the two sets contain the same items but in different order.

TABLE LVI

Summary of the Distribution of the Buchta Test When Given in Sets A and B

Set A	Set B
Mean..... 34.64 ± .61**	Mean..... 35.73 ± .59**
S. D..... 7.50	S. D..... 7.50
Difference 1.09 ± .85‡‡	
1.09 ÷ .85 = 1.3	

There is no significant difference in the means.

Table LVII shows similar distributions for the second objective test (laboratory and text test). The similarity of the means and standard deviations give evidence of the propriety of giving tests in this manner to eliminate cheating, for example.

TABLE LVII

Summary of the Distribution of the Second Objective Quiz When Given in Sets A and B

Set A	Set B
Mean..... 29.73 ± .61**	Mean..... 30.04 ± .58**
S. D..... 7.68	S. D..... 7.02
C. V..... .26	C. V..... .23
Difference in means .31 ± .84‡‡	
.31 ÷ .84 = .4	

There is no significant difference in the means.

Does the Manner of Scoring Tests Affect Students' Rank Rating?

It had originally been the thought of the writer to have the final objective examination scored with a credit of 1 for each blank

to be filled in. The questions originally submitted were arranged to make these blanks, when filled in, of approximately equal value as far as could be judged subjectively. For example, a blank "fill-in" would require something like a single judgment or item—not many items. If a formula were involved, only one quantity would be omitted, or a credit of 1 would be given for each item omitted. Likewise, on verbal questions one key-word omitted would receive a credit of 1.

While it is not possible at present to isolate what might be termed unit judgments, it seems wise to get as near such an eventuality as possible. A credit of 1 on a whole formula containing many quantities is obviously not the equivalent of a credit of 1 on each well-selected omitted item in a selected formula. Similarly, a credit of 1 on a sentence containing four or five omitted words is obviously not equivalent to a credit of 1 on each carefully selected word omitted.

In order to get data on differences in rank order due to the difference in scoring, 100 papers taken at random from the list of papers of students in engineering were rescored, giving credits to what might more properly be called units than whole questions. The students were ranked by each method of scoring. Table LVIII gives the resulting data.

TABLE LVIII

RANKING OF 100 STUDENTS IN THE FINAL OBJECTIVE TEST ON ORIGINAL AND REVISED SCORES

ORIGINAL RANK	REVISED RANK	ORIGINAL RANK	REVISED RANK
1	1	53	51
2	2	40	51
10	3	67	51
6	3	57	54
7	5	47	54
3	6	57	54
7	6	57	54
22	8	82	58
3	9	34	58
12	9	57	60
18	11	57	60

TABLE LVIII—*Continued*

RANKING OF 100 STUDENTS IN THE FINAL OBJECTIVE TEST ON ORIGINAL AND REVISED SCORES

ORIGINAL RANK	REVISED RANK	ORIGINAL RANK	REVISED RANK
7	11	57	60
14	11	67	60
22	14	53	60
22	14	65	65
22	14	53	65
3	14	67	65
14	14	57	65
18	19	77	69
10	19	40	69
30	19	76	69
47	22	67	72
22	22	65	72
18	22	67	74
16	25	67	74
22	25	47	74
12	27	80	77
34	27	67	77
16	27	80	79
18	27	83	79
51	27	91	79
28	32	77	82
40	32	75	82
34	34	77	84
30	34	88	85
47	34	88	85
28	34	83	87
32	34	83	88
30	39	91	88
34	39	91	88
34	39	87	91
30	42	83	91
40	43	88	93
53	43	94	94
34	45	96	94
40	45	96	96
67	47	94	97
51	47	98	98
57	49	98	98
40	49	100	100

The coefficient of correlation between these ranks was found to be .94±.01 which is high, but which allows for considerable individual variation. An examination of the ranks will disclose some serious variations. For example, ranks of 10 and 3 respectively; 22 and 8; 3 and 9; 12 and 27; 51 and 27; 34 and 45; 82 and 58; and 40 and 69.

Each of these sets of scores was correlated with the final numerical grade in mechanics scaled, giving values of .64±.04 for the original scores and .68±.03 for the revised scores. The method of giving credit on more nearly equal items (as blanks) seems in this case to be somewhat more valid, though there is not a great amount of difference. The difference is actually less than its probable error. The revised scoring (1 credit per blank) makes possible a greater score range and would seem to be preferable on this account alone.

Are These Tests Reliable, Valid, and Significant?

It was not advisable to repeat any of these tests for reliability computations. Therefore, the technique of getting scores on the even-numbered questions and the odd-numbered questions was resorted to. Table LIX gives some of the correlations between these sets of scores, together with the revised coefficient found by means of the Spearman-Brown formula.[1]

In using the Spearman-Brown formula it is assumed that the even-numbered and the odd-numbered questions, respectively, com-

TABLE LIX

Correlations between Halves of Achievement Tests in Mechanics

Test	Coefficient	P.E.	Revised Coefficient	P.E.	No. of Cases
1. Buchta test (Set A)	.58	.04	.73	.03	147
2. Buchta test (Set B)	.51	.04	.68	.03	139
3. Second objective test (Set A)	.70	.03	.82	.02	161
4. Second objective test (Set B)	.61	.03	.76	.02	151
5. Final objective test original scores	.75	.03	.86	.02	115
6. Final objective test revised scores	.80	.02	.89	.01	101

[1] Garret, *loc. cit.*

pose two forms of a test which should give comparable results. The reliability coefficient obtained by this method is probably a more severe test of reliability than one obtained by using the same test twice, so it may be expected that the reliability coefficients obtained are lower than those which would have been obtained by two administrations of the same test. The final objective test is, apparently, somewhat more reliable than either of the others, judged by the data given, and it is somewhat more reliable when a credit of 1 is given for each blank. Considering that it was a new and untried test, it has high reliability. Validity and significance are similar and not clearly distinguishable, perhaps. Validity refers to the content and seeks to determine whether or not test questions actually deal with informations or abilities associated with the sub-

TABLE LX

Correlations Showing Validity of Achievement Tests in Mechanics

Test	Score	Criterion	Coefficient	P.E.	No. of Cases
1. Buchta	M	Numerical grades, M score	.68	.02	300
2. Final objective	Raw ×2	Numerical grades, M score	.65	.02	301
3. Final objective	Original	Numerical grades, M score	.64	.02	98
4. Final objective	Revised	Numerical grades, M score	.68	.03	98
5. Final objective	Original	Buchta test, M score	.71	.02	305
6. Final objective and Buchta	Av. M	Numerical grades, M score	.73	.03	143

ject by authorities. Significance is more directly concerned with whether or not the test actually isolates the individual who is superior in the work of the course. "Do the questions sample the recognized content of the subject?" is the question which validity seeks to answer. "Do the scores gained rank the students in their correct order of achievement?" is the question which significance seeks to answer.

To test validity the questions of the test should be such as to be peculiarly characteristic of the subject content. Could one who had never had the course answer them as well as one who had taken the course? Or, can one who takes the course do better with them after, than before? The correlations given in Table LX furnish evidence of validity and significance.

The correlations give evidence that the tests are fairly valid considering the present status of marks and test ratings in general. As marks themselves are not very reliable, we cannot expect higher validity indices than those given. As objective devices supplant subjective opinions, higher indices will gradually be obtained.

It may be worth noting that the final conventional essay test correlates .66±.02 with the final grade. It is apparently nearly as good a single index as either objective test alone, but less so than the Buchta and final objective combined.

Tables LXI and LXII give further data showing the significances of the Buchta and final objective tests, respectively. In making interpretations it is well to remember that the scaled values are M values, which are directly comparable.

TABLE LXI

Comparisons of Individual Ratings in Final Grades and Scores on Buchta Test

			Per Cent
No. of ratings differing by 1 scale point or less	56	or	18.7
No. of ratings differing by 6 scale points or less	180	or	60.0
No. of ratings differing by 11 scale points or less	257	or	86.0
No. of ratings differing by more than 11 scale points	43	or	14.0

Coefficient = .68 ± .02. M scores used.
Number of ratings compared—300.

TABLE LXII

Comparisons of Individual Ratings in Final Grades and Scores on Final Objective Test

			Per Cent
No. of ratings differing by 1 scale point or less	29	or	9.5
No. of ratings differing by 6 scale points or less	198	or	64.9
No. of ratings differing by 11 scale points or less	267	or	87.5
No. of ratings differing by more than 11 scale points	38	or	12.5

Coefficient = .65 ± .02 M scores used.
Number of ratings compared—305.

Computing the multiple correlation Buchta test and final objective with the final grade as criterion gives a value of $R_{c.12} = .72$, which is about the same as the correlation, Number 6, in Table LX.

Do Students Prefer Objective Tests?

When the Buchta test was given, students were asked to state which type of test they preferred. Their opinions are given in Table LXIII.

TABLE LXIII

Students' Preference for the New Type of Test versus the Conventional Type

	Set A	Set B	Total
1. Prefer new type	96	83	179
2. Prefer conventional type	43	44	87
3. Do not say or have no preference	12	23	35

The preference is very much in favor of the new type of test. The Buchta test, it will be remembered, is a combination of multiple choice, matching, and uncompleted statements. More evidence on this point is given in a questionnaire summary in Chapter X.

Do Those Who Have Had High School Physics Obtain Higher Achievement Ratings in Mechanics?

The results show a significantly higher achievement by the students having had high school physics. Remembering that there was no significant difference in the means of the intelligence ratings in the two groups, it is apparently helpful to the average student in mechanics to have studied physics in high school. The differences, however, are not great as shown in Table LXIV. An

TABLE LXIV

Summary of Final Achievement Ratings in Mechanics, to Compare Those Who Have Had High School Physics With Those Who Have Not

Distributions	
H.S. Physics	Non-H.S. Physics
Mean..... 50.28 ± .64**	Mean..... 46.05 ± 1.23**
S. D..... 9.97	S. D..... 8.85
Difference in means 4.23 ± 1.39‡‡	
4.23 ÷ 1.39 = 3.04	

There is a significant difference in the means.

Univ. of Arizona Library

inspection of the original, complete table shows that some individuals, not having studied physics in high school, received higher marks than some who had studied it.

How Do the Students in the Various College Groups Compare in Achievement?

As these groups were compared in their initial ability in high school physics, it seemed worth while to compare them at the end of the course in achievement in mechanics. Tables LXV and LXVI make these comparisons. There are no significant differences in any of the means or S.D.'s. The chemistry group, however, shows a higher tendency. This difference would be considered significant by some authorities.

TABLE LXV

Summary of Engineering, Chemistry, and Academic Students in Achievement Ratings in Mechanics

	Number of Students		
	195	40	49
	Engineering	Chemistry	Academic
Mean	48.34±.66	52.0 ±1.50	48.21±1.43
S. D.	9.20±.47	9.50±1.05	10.00±1.01

M scores used.

Table LXVI makes direct, statistical comparisons.

TABLE LXVI

Comparisons of Achievement Ratings in Mechanics of Various College Groups

	Difference in Means	Difference ÷ S.E. of D.	Difference in S.D.'s	Difference ÷ S.E. of D.
1. Engineering and chemistry	3.66±1.64	2.2	.3±1.15	.3
2. Engineering and academic	.13±1.57	.1	.8±1.11	.7
3. Academic and chemistry	3.79±2.07	1.8	.5±1.46	.3

It will be remembered that in comparing these groups in intelligence the chemists were somewhat higher. This suggests a relationship between achievement and intelligence. Though the correlation coefficient was but .25, it will be noticed on examination of the correlation table (given in the next chapter) that it does, to a considerable extent, require a high degree of intelligence to get the high ratings. All individuals who received a percentage rating of 85 or above in mechanics had an intelligence rating of 48 or above. Only one had a rating of 48; the next lowest was 72. Table LXVII shows a comparison of those receiving a final mark of D or above with those receiving a mark less than D (including those who cancelled the course) in the freshman percentile ratings in intelligence. This furnishes further data on the effect of intelligence.

TABLE LXVII

SUMMARY OF DISTRIBUTIONS OF FRESHMAN PERCENTILE INTELLIGENCE RATINGS

THOSE D, OR ABOVE		THOSE BELOW D	
Mean	62.04 ± 2.27**	Mean	53.21 ± 2.39**
S. D	26.40	S. D	27.62
C. V	.42	C. V	.52
Difference in means		8.83 ± 3.30‡‡	
		8.83 ÷ 3.30 = 2.7	

99.7 chances in 100 of a true difference in the means greater than 0.

It is seen that the group of persons having marks of D, or above, had a higher mean percentile rating than the group having marks lower than D.

CHAPTER VIII

PAIRING STUDENTS FOR EDUCATIONAL EXPERIMENTS

The students majoring in architecture were not required to take the laboratory work. They, therefore, form a group with which another group might be equated for the purpose of studying the effects of the laboratory work. If the laboratory work produced effects, these effects might be distinguished in the achievement of those doing the laboratory work. The absence of its effects might also be discernible in those not doing the work. The plan of the study was to pair those students majoring in architecture with certain other students who were required to do laboratory work and make a contrasting study of the two groups thus formed, both as groups and as individuals within the groups.

In studies of this kind the students are usually divided into groups on the bases of certain adopted criteria for equation. In this study, the architectural group being fixed, it became necessary to find individuals in other groups who might be paired with those in architecture. In the absence of reliable and valid data on pairing students for a study of this kind, it was thought best to base the criteria for pairing on the correlations obtained between achievement in the course and other facts obtainable. This meant that no pairing could be done until the ratings for the course were completed at the end of the quarter.

Most studies of this kind assume that intelligence ratings and past school records are the best criteria for prognosis of future achievement. Enough work has been done amply to demonstrate this in a general way. That is, taking the school population in general, those in the upper half in intelligence ratings and in the upper half in past school records are much more likely to be in the upper half in future school achievement, as a whole. But in a group such as that in mechanics, it cannot be assumed without proof that those ranking highest in intelligence ratings and past school records will range themselves in a similar order in achievement in mechanics. In this study no such assumption was made. Rather, the past records which were available were gathered and compared with

TABLE LXVIII

CORRELATIONS OBTAINED BETWEEN FINAL ACHIEVEMENT RATINGS IN MECHANICS AND OTHER PAST RECORDS, RESPECTIVELY

PAST RATINGS	COEFFICIENT	P.E.	NO. OF CASES
Freshman intelligence percentiles	.25	.04	209
Pre-tests in H.S. physics, total raw score	.32	.04	237
First pre-test, only M scores	.36	.03	287
First pre-test, only M scores*	.45	.05	134
Honor points per credit	.53	.03	234
Chemistry (letter grades), M scores	.64	.03	201
Algebra and trigonometry (letter grades), M scores	.54	.03	236
Algebra and trigonometry (letter grades), M scores*	.57	.04	166
College algebra (letter grades), M scores	.43	.05	120
Shop (letter grades), M scores	.22	.05	179
Drawing (letter grades), M scores	.15	.05	207
English (letter grades), M scores	.34	.04	212
Multiple correlation mechanics (M scores) as criterion and intelligence and first pre-test	.44		

* Only sophomore engineering students who had taken high school physics and had not taken mechanics before were included.

final ratings in mechanics to determine, if possible, what relationship existed.

A common method of computing and expressing relationship is the correlation method. Table LXVIII shows the correlations obtained between achievement ratings in mechanics and each of the criteria obtainable from past records, respectively. These records included data on the following points: freshman intelligence percentile ratings; preliminary test ratings in high school physics; average honor points in past college work; and achievement in the former college courses of chemistry, freshman mathematics combined, college algebra alone (for engineering students), freshman English, engineering shop work, and engineering drawing.

It had been expected that the preliminary tests might be found to be valuable devices for prognosis of achievement in mechanics. The correlations obtained, as seen in Table LXVIII (.32, .36, and .45), apparently make them less prognostic than had been desired. It might be argued that the final numerical mark, as the achievement criterion, is perhaps not as valid or reliable as the scores from the objective tests would be. The Buchta and objective final tests,

combined, were therefore used as a criterion, the correlation between the first preliminary test and this criterion being found to be .48 ± .04, using only 134 sophomore engineering students who had taken high school physics and had not taken mechanics before. The M scores, on which many of the correlations are based, were obtained in the case of those for past subject ratings by getting distributions of those ratings from the Registrar's office and computing M scores, shown in Table LXIX. The M scores for the letter marks are used for correlations so as to have comparable ratings to correlate. A's, B's, C's, etc., probably do not mean the same as given by different instructors. The numerical values, which are based on the frequency of each mark given to members of the same group of students, are more likely to be comparable by the use of M scores than by any other available scheme occurring to the mind of the writer.

Even though the correlations in Table LXVIII are not as high as might be desired, they are positive and hence show that definite relationships exist. They make it possible to find critical ratings which may serve to differentiate those most likely to reach a certain degree of performance from those not likely to do so. For example, no person in the lowest ten percentiles in intelligence ratings received a higher mark than 50, which means that all students in the lowest ten percentiles failed. Again, no student in the highest ten percentiles in intelligence received a lower mark than 45. Only four of those received a lower mark than 60, which was the lowest

TABLE LXIX

TABLE TO SHOW SCALED M VALUES OF LETTER MARKS A, B, C, D, E, AND F, IN VARIOUS COURSES FOR STUDENTS ENROLLED IN MECHANICS

RATING	CHEMISTRY		DRAWING		ALGEBRA AND TRIGONOMETRY	
	Frequency	Scaled Value	Frequency	Scaled Value	Frequency	Scaled Value
E or F	76	33.8	48	32.0	124	36.0
D	222	43.5	149	41.0	287	46.1
C	253	52.2	311	50.7	231	54.8
B	150	60.9	111	60.2	88	62.3
A	25	71.4	47	68.0	40	69.5
	726		666		770	

TABLE LXIX—*Continued*

TABLE TO SHOW SCALED M VALUES OF LETTER MARKS A, B, C, D, E, AND F, IN VARIOUS COURSES FOR STUDENTS ENROLLED IN MECHANICS

RATING	SHOP		ENGLISH		COLLEGE ALGEBRA	
	Frequency	Scaled Value	Frequency	Scaled Value	Frequency	Scaled Value
E or F	8	25.3	69	33.5	52	38.0
D	174	40.2	263	44.4	76	47.3
C	243	50.5	319	55.4	64	55.3
B	127	59.8	33	67.4	27	62.9
A	32	69.2	12	73.8	9	70.5
	584		696		228	

D mark. Of those receiving a final grade of 85, or above, but one received less than 50 in intelligence. The correlation charts, which follow, give the graphic picture of the various relationships indicated. From these charts the critical ratings may be obtained, as illustrated above.

A study of the charts will indicate the need for some revision of the marking system in order to make such studies as these more worth while. More than a five-point scale is necessary. A fifteen-point scale is better and the test upon which the marks are based should be sufficiently well developed to spread students well along this fifteen-point scale.

WHAT DO THE CORRELATIONS REVEAL WHICH CAN BE USEFUL IN PAIRING THE STUDENTS?

Hull[1] summarizes the value of the coefficient of correlation as follows: below .50, practically useless for prognosis; .50-.60, possibly useful; .60-.70, genuine but limited value; .70-.80, of decided value but rarely found; above .80, not obtained by present methods. Judged by this classification, the only correlations that are worth any consideration for this study are those found between ratings in mechanics; and honor points per credit, chemistry, and mathematics (algebra and trigonometry), respectively.

[1] Clark L. Hull, "The Correlation Coefficient and Its Prognostic Significance," *Journal of Educational Research*, 15: 327-338 (March, 1927).

CORRELATION CHART I

NUMERICAL GRADES (X) AND INTELLIGENCE PERCENTILES (Y) IN MECHANICS

Y-VARIABLE From	To	20.0–24.9	25.0–29.9	30.0–34.9	35.0–39.9	40.0–44.9	45.0–49.9	50.0–54.9	55.0–59.9	60.0–64.9	65.0–69.9	70.0–74.9	75.0–79.9	80.0–84.9	85.0–89.9	90.0–94.9	95.0–99.9
96.0	101.5								1		2	2	1		1		
90.0	95.9						2	1	1	6	5	4	2	3	1		1
84.0	89.9				1	1	4	1	4	3	3	1	1	2			
78.0	83.9			1				2	3	1	2	3	3	1		2	
72.0	77.9	1			1			1		2	3	1	1	1	1		
66.0	71.9	1			1	1		2	4	1	2	3	1				
60.0	65.9				1	2			1	3	1	1	3				
54.0	59.9		1			1		1		6	4	1		1			
48.0	53.9					1		1	2	1	5				1		
42.0	47.9							1	2	2		1	3				
36.0	41.9			1	1	1		1	1	6	3	2	1				
30.0	35.9				3		1			7	1	1					
24.0	29.9				1		2	1	2	2							
18.0	23.9				1		1	2		1	2						
12.0	17.9									2	1	3	2				
6.0	11.9				1	1				1	2						
0.0	5.9		1		1		1	1									

X-VARIABLE

Coefficient = .25 ± .04.

CORRELATION CHART II

NUMERICAL GRADES (X) AND PRELIMINARY TEST—TOTAL RAW SCORE (Y) IN MECHANICS

Y-VARIABLE From To																	
150.0–159.9															1		1
140.0–149.9													2				
130.0–139.9										1	2						
120.0–129.9														1	1		
110.0–119.9								1		2	3	2	5	4		1	
100.0–109.9							1	1	1	6	5	5	3		1	1	
90.0– 99.9				1	2		3	2	2	5	7	5	2	1			
80.0– 89.9							1	1	3	2	8	4	2	1	2		
70.0– 79.9			1	1	2	2	2	5	6	5	3	3	1			1	
60.0– 69.9		1	1		2			2	3	9	6	3	3			1	
50.0– 59.9	1				2	2		1	3	8	2	3	1				
40.0– 49.9					1	1		1	4	4	5	1	1	1			
30.0– 39.9						1	1	1	1	3	1		3				
20.0– 29.9					1	2	2		3	1		1	1				
10.0– 19.9									3			1					
0.0– 9.9													1				
X-VARIABLE From	15.0	20.0	25.0	30.0	35.0	40.0	45.0	50.0	55.0	60.0	65.0	70.0	75.0	80.0	85.0	90.0	95.0
X-VARIABLE To	19.9	24.9	29.9	34.9	39.9	44.9	49.9	54.9	59.9	64.9	69.9	74.9	79.9	84.9	89.9	94.9	99.9

Coefficient = .32 ± .04.

CORRELATION CHART III

NUMERICAL MARKS (M SCORES) (Y) AND PRELIMINARY FIRST TEST (M SCORES) (X) IN MECHANICS

Y-VARIABLE From To																	
79.0–82.9																	1
75.0–78.9													1	1			
71.0–74.9							1									1	
67.0–70.9								1		2		1			1		
63.0–66.9							1		1	2	2	1		2			
59.0–62.9		1			1	2	6	1		6	3	1	3	2			
55.0–58.9				1	1		4	4	7	9	4	3	2				
51.0–54.9				1	2	3	7	7	7	10	2	6	1	2			
47.0–50.9					2	6	11	6	11	8	6	1	1	1			
43.0–46.9				3	4	2	8	9	2	7							
39.0–42.9		1		4		3	10	4	3	6	2	2	4				
35.0–38.9					4	1	1	2	2	1							
31.0–34.9		1					8	2	3	1							
27.0–30.9					1			3	1								
23.0–26.9								1									
19.0–22.9								1									
15.0–18.9																	
X-VARIABLE From	15.0	19.0	23.0	27.0	31.0	35.0	39.0	43.0	47.0	51.0	55.0	59.0	63.0	67.0	71.0	75.0	79.0
X-VARIABLE To	18.9	22.9	26.9	30.9	34.9	38.9	42.9	46.9	50.9	54.9	58.9	62.9	66.9	70.9	74.9	78.9	82.9

Coefficient = .86 ± .03.

CORRELATION CHART IV

NUMERICAL GRADES (M SCORES) (X) AND FIRST PRELIMINARY TEST (M SCORES) (Y) IN MECHANICS

(Only sophomore engineering students who have had high school physics and who have not had mechanics before, included)

Y-VARIABLE From To																			
78.0–80.9																			1
75.0–77.9																	1		
72.0–74.9												1							
69.0–71.9												1				1			
66.0–68.9								2					1		1			1	
63.0–65.9								1		1		4	3						
60.0–62.9							1	1		4		3	2		2		1		
57.0–59.9									5	1		2	3		4				
54.0–56.9						1		1	3	1	3	4	4		1	1			
51.0–53.9					1	1	1	1		1	7	6	5						
48.0–50.9			1	1			1		2	7	6	4							
45.0–47.9						5			3	1	2	1	1						
42.0–44.9						1	1	2		1	1	2			1				
39.0–41.9											1	1							
36.0–38.9																			
33.0–35.9									1			1							
30.0–32.9									1										
X-VARIABLE From	20.0	23.0	26.0	29.0	32.0	35.0	38.0	41.0	44.0	47.0	50.0	53.0	56.0	59.0	62.0	65.0	68.0	71.0	74.0
X-VARIABLE To	22.9	25.9	28.9	31.9	34.9	37.9	40.9	43.9	46.9	49.9	52.9	55.9	58.9	61.9	64.9	67.9	70.9	73.9	76.9

Coefficient = .45 ± .05.

CORRELATION CHART V

NUMERICAL GRADES (M SCORES) (x) AND HONOR POINTS PER CREDIT (y) IN MECHANICS

Y-VARIABLE From To	15.0 19.0	19.0 22.9	23.0 26.9	27.0 30.9	31.0 34.9	35.0 38.9	39.0 42.9	43.0 46.9	47.0 50.9	51.0 54.9	55.0 58.9	59.0 62.9	63.0 66.9	67.0 70.9	71.0 74.9	75.0 78.9	79.0 82.9	83.0 86.9
270.0–284.9																1		
255.0–269.9												1						
240.0–254.9																		
225.0–239.9												1	1		1			
210.0–224.9										1	1		1					
195.0–209.9								1	1		1			1				
180.0–194.9									3			2	3					
165.0–179.9						1	2	1	1	3		4		1				
150.0–164.9									2	2	3	1						
135.0–149.9							1	1	1	3	3	4						
120.0–134.9				1			1	2	4	8	5	1	2					
105.0–119.9				1	1	1	6	5	7	8	5	1						
90.0–104.9					1		4	6	9	5	4	3	1					
75.0– 89.9				1	3		6	4	8	7	4							
60.0– 74.9		1			3	1	9	3	6	3	1							
45.0– 59.9				1		1	6	4	6	2	1							
30.0– 44.9					2	1		3	1									
15.0– 29.9																		
X-VARIABLE From	15.0	19.0	23.0	27.0	31.0	35.0	39.0	43.0	47.0	51.0	55.0	59.0	63.0	67.0	71.0	75.0	79.0	83.0
To	18.9	22.9	26.9	30.9	34.9	38.9	42.9	46.9	50.9	54.9	58.9	62.9	66.9	70.9	74.9	78.9	82.9	86.9

Coefficient = .53 ± .03.

CORRELATION CHART VI

NUMERICAL GRADES (M SCORES) (x) AND CHEMISTRY LETTER GRADES (M SCORES) (y) IN MECHANICS

Y-VARIABLE From To																		
83.0–86.9																		
79.0–82.9																		
75.0–78.9																		
71.0–74.9														1			1	
67.0–70.9														1	1	1		
63.0–66.9									2		2	3		1				
59.0–62.9							1			5	2	3	2	1				
55.0–58.9							1	4	10	10	8	5	1					
51.0–54.9							5	3	7	6	6	2	1					
47.0–50.9				2	1	1	3	5	9	10	2	2						
43.0–46.9		1			4		15	12	10	6	4							
39.0–42.9					3	1	2	2	4		1							
35.0–38.9				1	1	1												
31.0–34.9						1												
27.0–30.9						1												
23.0–26.9																		
X-VARIABLE From	15.0	19.0	23.0	27.0	31.0	35.0	39.0	43.0	47.0	51.0	55.0	59.0	63.0	67.0	71.0	75.0	79.0	
X-VARIABLE To	18.9	22.9	26.9	30.9	34.9	38.9	42.9	46.9	50.9	54.9	58.9	62.9	66.9	70.9	74.9	78.9	82.9	

Coefficient $=.64 \pm .03$.

CORRELATION CHART VII

NUMERICAL GRADES (M SCORES) (x) AND FRESHMAN MATHEMATICS (ALGEBRA AND TRIGONOMETRY) LETTER MARKS (M SCORES) (y) IN MECHANICS

Y-VARIABLE From	To	4.0	8.0	12.0	16.0	20.0	24.0	28.0	32.0	36.0	40.0	44.0	48.0	52.0	56.0	60.0	64.0	68.0	72.0
80.0	82.9																		
77.0	79.9																		
74.0	76.9																		
71.0	73.9																		
68.0	70.9													1			3		1
65.0	67.9													1	1	2	2	1	
62.0	64.9									1			3	2	1	3			1
59.0	61.9								1				1	4	2	2			
56.0	58.9									1		3	2	6	5	2			
53.0	55.9									2	2	6	10	6	4	5	2	1	
50.0	52.9								1	2	5	5	13	12	6	2	1		
47.0	49.9								1	1	3	1	6	3	1	1			
44.0	46.9						1	2	5	7	2	10	12	6	5				
41.0	43.9					1			4	5	6	7	4	1	2				
38.0	40.9									1									
35.0	37.9								1										
X-VARIABLE	From	4.0	8.0	12.0	16.0	20.0	24.0	28.0	32.0	36.0	40.0	44.0	48.0	52.0	56.0	60.0	64.0	68.0	72.0
	To	7.9	11.9	15.9	19.9	23.9	27.9	31.9	35.9	39.9	43.9	47.9	51.9	55.9	59.9	63.9	67.9	71.9	75.9

Coefficient = .54 ± .03.

CORRELATION CHART VIII

NUMERICAL GRADES (M SCORES) (x) AND FRESHMAN MATHEMATICS (ALGEBRA AND TRIGONOMETRY) LETTER MARKS (M SCORES) (y) IN MECHANICS

(Only sophomore engineers having had high school physics and not having had mechanics before, included)

Y-VARIABLE From To																		
69.0–70.9																	1	
67.0–68.9														1				
65.0–66.9											1			1	1	1		
63.0–64.9										2								
61.0–62.9											4	2		2				1
59.0–60.9										1	2	1		1				
57.0–58.9					1				3	1	4	3						
55.0–56.9										1				1				
53.0–54.9						1		2	3	8	6	5		4	1	1		
51.0–52.9								1	3	5	2			1				
49.0–50.9					2	1	1	4	1	6	11	5		1				
47.0–48.9								3	1		1			1				
45.0–46.9			1	2	3		3	1	7	4	5	2						
43.0–44.9							1	2				1						
41.0–42.9	1				3	1	1	3	5	2								
39.0–40.9																		
37.0–38.9																		
35.0–36.9					1													
X-VARIABLE From	20.0	23.0	26.0	29.0	32.0	35.0	38.0	41.0	44.0	47.0	50.0	53.0	56.0	59.0	62.0	65.0	68.0	71.0	74.0
X-VARIABLE To	22.9	25.9	28.9	31.9	34.9	37.9	40.9	43.9	46.9	49.9	52.9	55.9	58.9	61.9	64.9	67.9	70.9	73.9	76.9

Coefficient $= .57 \pm .04$.

CORRELATION CHART IX

NUMERICAL GRADES (M SCORES) (X) AND FRESHMAN ENGLISH LETTER MARKS (M SCORES) (Y) IN MECHANICS

Y-VARIABLE From To																		
75.0–77.9																		
72.0–74.9																1		
69.0–71.9							1							1				
66.0–68.9									1	3		1	2					1
63.0–65.9					2		1	2		5		3	4		3	1		1
60.0–62.9												1				1		
57.0–59.9								2	4			7	6		3			
54.0–56.9		1		2			1	1	5	5		4	4		3	1	2	
51.0–53.9							3	4	8	7		7	4		4	1		
48.0–50.9							3	3	3	10		9	2		2			
45.0–47.9															1			
42.0–44.9			1		4	2		3	3	6		7	4		1			
39.0–41.9							1	1	4	4		1						
36.0–38.9					3	2		2	2	1								
33.0–35.9							1		1	1								
X-VARIABLE From	20.0	23.0	26.0	29.0	32.0	35.0	38.0	41.0	44.0	47.0	50.0	53.0	56.0	59.0	62.0	65.0	68.0	71.0
X-VARIABLE To	22.9	25.9	28.9	31.9	34.9	37.9	40.9	43.9	46.9	49.9	52.9	55.9	58.9	61.9	64.9	67.9	70.9	73.9

Coefficient = .34 ± .04.

These findings put in question all experimental studies in which students have been paired for equation of groups. Especially do they make it worthless to consider any one student as comparable with another, unless achievements over a considerable period are compared and found identical in a great variety of activities. If groups, however, show equal mean achievement records in many school activities, are of the same variability, and then suddenly show a significant difference in achievement in a given activity, it is certainly worth while to try to discover the reasons for such a difference. If the architectural group is found to be practically identical with some other group, or groups, in achievement in a great variety of activities and is then found to show a significant difference in achievement in mechanics, the reasons may be due to the most notable difference in their treatment, viz., the lack of the two-hour laboratory period.

CHAPTER IX

GROUP COMPARISONS IN MECHANICS

The Architectural Group

On the original list of students majoring in architecture were twenty-eight persons. One was a woman; thirteen men cancelled the course during the quarter; fourteen received a final grade. One was in the hospital during the final examinations, but six were left for whom complete records are available. None of the students in architecture had taken chemistry in college. Ten had complete records, except for ratings in English and intelligence. Table LXX gives the list of the ten students with fairly complete records. Ratings in mathematics, first pre-test, Buchta, final objective, and second objective tests, and the final grades in mechanics, are all given in M scores.

It was thought most worth while to use a group of architectural engineering students to compare with the group listed above. The architectural engineering students had the same work during the first year and were more comparable with the architectural group, apparently, than any other available group.

The only ratings giving correlations above .50 with achievement in mechanics were those in mathematics and average honor points. Hence, these were used as the bases of pairings. The ratings for the first pre-test are included in the tables but were not used in pairing. Table LXXI presents the data. It will be noticed that there are discrepancies in the pairs even in the bases used, but an attempt was made to balance a lower rating in one criterion with a higher in the other. The two members paired were picked so that each had the same status with reference to high school physics and a previous course in mechanics.

The architectural engineering group was apparently a somewhat better group than the architects, judged by ratings in mathematics and average honor points. Six of the group exceeded their paired companions in architecture in pre-test ratings as well. The record of number 10 in the pre-test is incomplete. One page of his test

TABLE LXX

Records of Architectural Students

Student	Math.	Av. H.P.	Had High School Physics	Had Mechanics Before	Tests					Final Grade
					First Pre-Test	Laboratory	Buchta	Final Objective	Second Objective	
1	50.4	.87	Yes	No	55.8	16	53.5	47.8	59.4	53.8
2	46.1	.43	Yes	No	46.5	15.5	44.5	42.7	53.9	49.2
3	43.2	.77	Yes	No	40.4	9	40.8	37.5	45.7	45.1
4	46.1	.67	Yes	No	46.5	8	26.0	31.2	33.2	33.7
5	42.7	1.27	Yes	No	46.5	12	37.4	33.4	48.5	45.1
6	50.4	.81	Yes	Yes	40.4	10	37.4	42.7	45.7	49.2
7	45.6	1.18	Yes	No	43.4	4	49.0	25.1	26.0	39.5
8	59.8	.63	Yes	No	52.8	10	58.8	61.5	51.2	58.0
9	51.9	.76	No	No	24.5	8	33.1	37.5	40.9	42.0
10	54.2	1.71	Yes	No	70.8	11	61.4	67.7	51.2	66.2
Mean	49.04± 1.59**	.91± .11**				10.35 ±1.08**	44.19± 3.43**	42.71 ±4.01**	45.57 ±2.99**	48.18 ±2.82**
S. D.	5.09	.35				3.39	10.85	12.59	9.44	8.93
C. V.	.10	.38				.33	.25	.29	.20	.19

TABLE LXXI

RECORDS OF PAIRED STUDENTS IN ARCHITECTURAL ENGINEERING FOR COMPARISON WITH ARCHITECTURAL STUDENTS

STUDENT	MATH.	AV. H.P.	HAD HIGH SCHOOL PHYSICS	HAD MECHANICS BEFORE	TESTS					FINAL GRADE
					First Pre-Test	Laboratory	Buchta	Final Objective	Second Objective	
1	50.4	.90	Yes	No	55.8	14	49.0	39.9	51.2	58.0
2	42.7	.56	Yes	No	49.3	21	40.8	35.3	65.5	42.0
3	54.8	1.27	Yes	No	66.7	10	40.8	35.3	45.7	42.0
4	41.0	.55	Yes	No	46.5	21	49.0	47.8	69.2	42.0
5	51.9	1.14	Yes	No	58.2	21	58.8	61.5	65.5	62.0
6	43.3	.69	Yes	Yes	52.8	14	58.8	61.5	53.9	58.0
7	58.5	1.12	Yes	No	55.8	20	61.4	45.2	65.5	53.8
8	54.8	1.49	Yes	No	52.8	17	44.5	45.2	48.5	58.0
9	51.5	.84	No	No	34.6	17	49.0	39.9	59.4	45.1
10	64.7	1.14	Yes	No		19	44.5	51.0	65.5	49.2
Mean	51.36± 2.24**	.97± .09**				17.40 ±1.13**	49.66 ±2.27**	46.26 ±2.85**	58.99 ±2.54**	51.01 ±2.36**
S. D.	7.08	.26				3.56	7.19	9.02	8.04	7.46
C. V.	.14	.27				.20	.14	.19	.14	.15

Paired students are numbered to correspond with those in Table LXX.

was missing. The score on the pages returned is 15 and the comparable score for his companion is 38.

Table LXXII (a) is given to show the statistical significance of the differences in the means of the ratings for the two groups. It merely attempts to answer the question as to whether or not, considering the small numbers in the groups, the differences are great enough to warrant the generalization that such differences would also be obtained (1) for larger groups if more students were included on the same basis; or (2) for other simpler samplings of small groups.

There are significant differences only in the laboratory test and in the second objective test. The laboratory test was a portion of the second objective test. The chances of there being a real difference in the laboratory ratings greater than 0 are 99.9 in 100 and in the second objective test 99.8 in 100.

Probably Table LXXII (b) gives the simplest manner of comparing this group with the group in architecture.

The results seem to indicate quite clearly that the architectural students have lower ratings on specific laboratory work, both when considered as a group or as individuals. There is a tendency for group ratings to be lower in other criteria, but it is not so clearly evident. It is rather dangerous, however, to draw conclusions from the data presented, in view of the fact that there are so many other factors which might affect the results that have not been deter-

TABLE LXXII (a)

COMPARISONS TO SHOW THE SIGNIFICANCE OF DIFFERENCES IN THE MEANS OF THE RATINGS OF ARCHITECTURAL AND ARCHITECTURAL ENGINEERING STUDENTS

	MATH.	AV. H.P.	TESTS				FINAL GRADE
			Lab.	Buchta	Final Obj.	Sec. Obj.	
Difference in means	2.22±2.75	.06±.14	7.05±1.56	5.47±4.11	3.55±4.92	13.42±3.92	2.83±3.68
Difference ÷ S.E. of difference	.81	.36	4.5	1.3	.7	3.4	.8

TABLE LXXII (b)

NUMBER AND PER CENT OF STUDENTS IN ARCHITECTURAL ENGINEERING EXCEEDING THEIR PAIRED COMPANIONS IN ARCHITECTURE

	NUMBER	PER CENT
Mathematics	4	40
Average honor points	5	50
First pre-test	6	60
Laboratory test	9	90
Buchta test	5	50
Final objective test	5	50
Second objective test	8	80
Final grade	6	60

mined. The groups are too small and the bases of pairing are too inadequate.

It was still felt to be worth while to select students from the other college groups on similar bases for purposes of comparison. Tables LXXIII and LXXIV are given for this purpose. The individual records, originally included, are here omitted. Pairings were made on exactly the same bases as in Tables LXX and LXXI.

The comparisons made give evidence that the laboratory test scores rate the members of all the groups higher than those of the group of architects given. That the cause of this difference can be assigned to the lack of a two-hour period of laboratory work a week may seem doubtful but seems reasonable when the lesser differences in the other ratings are considered.

The mean of the laboratory ratings to all the six groups, other than the architectural group, is 15.10. But two of the ten members of the architectural group studied exceeded this mean and these two but slightly. The mean of the Buchta test ratings in the six groups is likewise 49.69. Three of the architectural students exceeded this. The mean of the final objective test ratings is 49.91. Two of the architectural students exceeded this. The mean of the final numerical grades for the six groups is 49.69. Three of the architectural students exceeded this.

The standard deviations given show the groups less variable than the whole class in most of the ratings. It will be remembered that in all ratings given except those in average honor points and

TABLE LXXIII

SUMMARY OF RECORDS OF PAIRED STUDENTS FOR PURPOSES OF COMPARISON

		Mathematics	Average Honor Points	Tests				Final Grade
				Laboratory	Buchta	Final Objective	Second Objective	
Architectural students	Mean	49.04	.91	10.35	44.19	42.71	45.57	48.18
	S. D	5.09	.35	3.39	10.85	12.59	9.44	8.93
	C. V	.10	.38	.33	.25	.29	.20	.19
Architectural engineering students	Mean	51.36	.97	17.40	49.66	46.26	58.99	51.01
	S. D	7.08	.30	3.56	7.19	9.02	8.04	7.46
	C. V	.14	.31	.20	.14	.20	.14	.15
Mechanical engineering students	Mean	49.49	.91	15.25	48.27	51.55	50.78	51.08
	S. D	3.36	.35	3.30	9.21	9.15	7.18	6.23
	C. V	.07	.38	.22	.19	.18	.14	.12
Electrical engineering students	Mean	49.86	.91	15.90	50.59	50.91	54.08	48.22
	S. D	4.55	.30	3.75	5.43	5.94	7.71	7.19
	C. V	.09	.33	.24	.11	.12	.14	.15
Civil engineering students	Mean	51.53	.84	14.70	48.43	50.20	49.06	48.46
	S. D	5.97	.30	2.37	6.62	3.99	4.83	2.08
	C. V	.12	.36	.16	.14	.08	.10	.04
Chemistry students	Mean	47.74	.99	13.55	48.19	52.77	49.44	52.14
	S. D	3.30	.34	2.34	8.39	9.62	6.18	8.03
	C. V	.07	.34	.17	.17	.18	.13	.15
Academic students	Mean	51.91	.85	13.80	53.01	47.78	48.48	47.23
	S. D	5.64	.33	4.54	8.54	4.04	8.84	3.80
	C. V	.11	.39	.33	.16	.08	.18	.08

TABLE LXXIV

PERCENTAGES OF STUDENTS IN EACH GROUP EXCEEDING THEIR PAIRED COMPANIONS IN ARCHITECTURE

	MATHEMATICS	AVERAGE HONOR POINTS	TESTS					FINAL GRADE
			First Pre-Test	Laboratory	Buchta	Final Objective	Second Objective	
Mechanical engineering	30	40	80	90	60	70	70	60
Electrical engineering	30	40	70	80	60	80	80	50
Civil engineering	50	30	60	90	50	70	50	40
Chemistry	10	50	80	80	70	80	60	50
Academic	50	30	30	80	70	70	50	50
Architectural engineering	40	50	60	90	50	50	80	60

in the laboratory test, M scores are used, 10 being the standard deviation of these scores. In most of the ratings in mechanics, however, the architectural group has standard deviations greater than 10, or nearly equal to it. This shows that the architectural group is a representative sample of the class in mechanics, but the other groups are not such good samples.

It is unfortunate that a larger number of architects did not complete the course. The comparisons made must be interpreted with a consideration of the small number of architects involved. It will be noticed that all of the groups have fewer members exceeding the architects in mathematics and average honor points, but more exceeding them in practically all of the other ratings. The evidence indicates, in general, that the architects rate lower in achievement in mechanics and that the reason is partly, at least, due to the lack of the laboratory work. Two considerations, however, must not be forgotten. First, the bases of pairing the students are not of thoroughly proven validity. Second, the small number of architects left makes the usual statistical treatment unsafe for drawing indisputable conclusions.

The apparently lower ratings of the architectural group in mechanics, especially in laboratory work, may be partly due to their lack of the two hours a week in the laboratory. This would be easier to believe were it not for the apparently lower ratings of the architects in the first pre-test. The only lower group was the academic group. This gives evidence of a possible lack of interest in physics on the part of the architects, as well as of the academic students. If interest is lacking, motivation for effort is lacking. This may be a vital reason for the low achievement in mechanics.

Comparisons of Students in Architecture with Those in Other Groups for Classes of 1923-27

Comparisons of the ratings of students in architecture who are not required to do the laboratory work with those of other groups who do laboratory work were also made for classes of former years. In order to make such comparisons the mean and sigma ratings and coefficients of variability of all architectural students from 1923-27, inclusive, were computed. These ratings were based al-

TABLE LXXV (a)

SUMMARY OF THE DISTRIBUTION OF FINAL NUMERICAL MARKS OF STUDENTS IN MECHANICS, YEARS 1923-27, INCLUSIVE

	ARCHITECTURE	ARCH. ENG.	ENGINEERING	ACADEMIC
Mean	64.21±1.39	64.88±2.15	68.00±1.69	68.50±1.61
S. D	11.88	12.51	14.34	13.77
C. V	.19	.19	.21	.20

TABLE LXXV (b)

RELIABILITIES OF DIFFERENCES BETWEEN MEAN RATINGS OF VARIOUS GROUPS IN MECHANICS, 1923-27, AND OF STUDENTS IN ARCHITECTURE

	72 ENGINEERING STUDENTS	72 ACADEMIC STUDENTS	34 ARCHITECTURAL ENGINEERING STUDENTS
Difference in means	3.79±2.19**	4.29±2.13**	.67±2.56**
Difference ÷ S. E. of difference	1.7	2.0	.3

most entirely on weekly written-quiz marks and quarter examinations.

Table LXXV gives, for purposes of comparison, the distributions of such ratings for seventy-two architectural students, seventy-two engineering students (exclusive of architectural engineers), seventy-two academic students, and thirty-four architectural engineering students enrolled during these years. There were not enough students in architectural engineering to make up the full quota of seventy-two. An equal number of students in each group was drawn from each year's class so as to discount any possible variations in the standards of marking from year to year.

As a group, the engineers and academic students rank somewhat higher than the architects. The architectural engineers rank practically the same. The evidence seems to indicate some factor tending to produce a tendency for lower achievement on the part of both architects and architectural engineers, though the former have no laboratory work, and the latter have. This factor may be one of interest or motivation or dependent upon some condition existing in the School of Architecture.

The lack of the laboratory work does not seem to produce a significant difference in achievement, however, between architects and architectural engineers as measured by the ordinary weekly quizzes or quarter examinations upon which the marks are based. It must be remembered, however, that quarter marks are given for the laboratory work separately from the regular marks for the lecture and quiz work. The architectural students do not receive these marks, of course, for they have no laboratory work. The fact, that in the consideration of the class of 1927-28 the architects do less well in the laboratory test proper, probably shows specific benefits from the laboratory work, not included in the quizzes or regular final examinations.

CHAPTER X

STUDENT OPINION IN MECHANICS

Student thought and attitude should be given at least some weight in the evaluation of college courses. If achievement is dependent upon motivation, and emotions are the drives which impel to achievement, it is clearly important to develop right emotional attitudes. In order to do this it is necessary to know present attitudes. Experiences are not the same for different individuals. The purposes of all students enrolling in a certain course are not identical. Instructors may have had different experiences and different purposes than students taking their courses. Or they may have forgotten their student days in some measure. In other words, they may be unable to see the viewpoints of the individuals in their classes. If a real cross-section of student opinion can be obtained, it may serve to make instructional problems clearer and eventually help to evolve better methods of instruction.

To get a cross-section of student opinion, questionnaires were sent to 353 students who had taken the fall course in mechanics. By February 9, 1928, seventy questionnaires had been returned. Two were from students who received a mark of A, four from B students, fourteen from C students, nineteen from D students, twenty-six from E, F, and cancelled students, and five from students of unknown status. In the following summary these are listed separately to note any possible variations in the responses of individuals of these different groups.

Questionnaire Summary III

QUESTIONNAIRE TO STUDENTS IN MECHANICS, FALL, 1927

1. Do you think this quarter's course in mechanics could have been made more practical for you?

	A's	B's	C's	D's	E's, F's, Cancel.	Others	Total
a. Yes		1	8	17	21	3	50
b. No		2	1		1	1	5
c. Not intended to be	1						1
d. Slightly	1		1		1		3
e. No answer		1	4	2	3		10
f. I don't know						1	1
Total							70

2. In what ways could the course have been made more valuable to you individually?

	A's	B's	C's	D's	E's, F's, Cancel.	Others	Total
a. Text more explanatory		1	4	2	6	1	14
b. Less theoretical and mathematical—more discussion of practical			2	5	8	1	16
c. Give opportunity for questions and discussions			1	1	1	2	5
d. Less stress on quizzes and marks						1	1
e. More illustrative problems or applications			4	4	2	1	11
f. Quite satisfactory						1	1
g. Teach in College of Engineering so as to be able to apply to specific problems					1		1
h. Instructor should be more interesting				3	4		7
i. Make classes smaller			1		2		3
j. Different instructor					2		2
k. Different methods					1		1
l. Lectures should not be duplicated in text		1	3	3	1		8
m. Waste less time in trivial experiments					1		1
n. No answer	1		2	1	2		6
o. Eliminate non-essentials			1		2		3
p. Make quizzes easier					1		1
q. Separate courses for different groups				1	1		2
r. Cover less ground					1		1
s. More laboratory work	1		1		2		4
t. More individual instruction			1		1		2
u. More understandable lectures					1		1
v. Eliminate final examination					1		1
w. Less memorization of formulae				1			1
x. Lectures louder				1			1
y. More explanation			2	5			7
z. Quiz instead of written homework			1				1
aa. Let students examine all quizzes		1					1
bb. More demonstration		1					1

3. Enumerate the specific values from the laboratory work in physics, which you think would not have been gained in any other way.

	A's	B's	C's	D's	E's, F's, Cancel.	Others	Total
a. Contact with physical apparatus	1						1
b. Knowledge of scientific apparatus			2	7	4	2	15
c. Skill in manipulation			2	1	1	1	5
d. Care in handling apparatus				1			1
e. Laboratory technique						2	2
f. Practice in physical measurement	1				1		2
g. Accuracy in measuring		1	1	2	1		5
h. Insight into physical measurement				1			1
i. Concepts of error in measures	1		2	2	1	1	7
j. Methods of experiment		1	2	2		1	6
k. Illustration				3	5	1	9
l. Shows applications of theory		1			4	2	7
m. Clarify principles		2	1	2	6		11
n. Fix principles in mind				1	1		2
o. Verification of principles				1	1		2
p. Understanding of terms				1			1
q. Learn to follow directions		1					1
r. Accuracy in observation			2		2		4
s. Patience			1				1
t. Resourcefulness			1				1
u. Ingenuity			1				1
v. Neatness			2		2		4
w. Accuracy			3	3	8	1	15
x. Mental exercise			1		1		2
y. Training in methodical thinking				1			1
z. Speed				1	1		2
aa. Carefulness					1		1
bb. Deeper interest in the subject					1		1
cc. Practice in addition and subtraction					1		1
dd. Better use of formulae			1				1
ee. Writing reports		1	1	1			3
ff. Of no value		1	1	3	2		7

4. From which of the following sources do you feel you received the greatest benefit: (1) lecture, (2) textbook study, (3) outside reading of reference books, (4) laboratory, (5) quiz, (6) discussion?

	A's	B's	C's	D's	E's, F's, Cancel.	Others	Total
Lecture	1	1	4	4	3	1	14
Textbook study		2	4	3	6	1	16
Reading (outside)			1	1	2		4
Laboratory			1	4	5	1	11
Quiz	1	1		1	1		4
Discussion (outside)			2	4	7	2	15
Demonstration				1			1
							65

5. List the others in (4) in order of importance to you (summary of only the last mentioned, or least important).

	A's	B's	C's	D's	E's, F's, Cancel.	Others	Total
1.			2	6	7		15
2.					2		2
3.	2	2	2	6	6	2	20
4.		1	2	3			6
5.			3		4	3	10
6.		1	1	2	2		6
							59

6. Would you elect physics if it were not prerequisite in your list of studies?

	A's	B's	C's	D's	E's, F's, Cancel.	Others	Total
Yes	2	4	10	12	15	2	45
No			4	6	11	1	22
Possibly				1			1
Don't know						2	2
							70

7. Specify your reason for answer to (6).

	A's	B's	C's	D's	E's, F's, Cancel.	Others	Total
a. Gives understanding of life phenomena	1	1	1	4	1		8
b. Very interesting, *or,* I like it	1	2	3	1	9	1	17
c. Good training	1	1					2
d. Makes one think, concentrate, reason		1					1

e. Practical	1	1		1	1	4
f. Fundamental to engineering		3	2	2		7
g. Too difficult		1	1	1	1	4
h. Fundamental to scientific work		1	1			2
i. Helps in chemistry		1				1
j. Broadens the mind		1		1		2
k. Uninteresting, *or*, I don't like it		1	3	1		5
l. Doesn't repay effort put upon it		1		1		2
m. No value to me			3	5		8
n. Dislike mathematics			1			1
o. Beneficial because difficult			1			1
p. Part of a liberal education			1	3		4
q. No value in medicine				2	1	3
r. Poor instruction				1		1
s. Not practical					1	1

8. Which type of test do you prefer in physics, the many-question, short-answer type, or the conventional essay or problem type?

	A's	B's	C's	D's	E's, F's, Cancel.	Others	Total
a. Many-question, short-answer	2	1	8	10	16	1	38
b. Conventional		3	4	6	8	2	23
c. No choice			1		1	1	3
d. A combination			1	3	1	1	6

9. Specify your reasons for your answer to (8).

	A's	B's	C's	D's	E's, F's, Cancel.	Others	Total
a. Conventional more natural				1		1	2
b. Conventional fairest				1		1	2
c. First covers more ground (more chance)	2	1	3	5	15	1	27
d. First more searching	1			1	1	1	4
e. First for general knowledge			4	4	4	1	13
f. Second for application			2	1	2	1	6
g. First may be memorized only					1		1
h. Second requires more thought		1	1	1	1		4
i. Second shows knowledge of subject		1	2		1		4
j. Second is shorter—more time to think		1			1		2
k. Second is easier				1			1
l. Second because mathematical				1			1
m. Problems not so catchy			1				1
n. First easier to tell what you know			1				1

o. First more guessing		1	1			2
p. First gives me better marks			1			1
q. First more understandable			1			1
r. Second covers only principles		1				1
s. Second covers only essentials		1				1

10. Do you think your final mark was fair?

	A's	B's	C's	D's	E's, F's, Cancel.	Others	Total
Yes	2	4	8	9	19	3	45
No			4	7	4	1	16
Question			2	3	3	1	9
Total							70

11. Specify your reasons for your answer to (10).

	A's	B's	C's	D's	E's, F's, Cancel.	Others	Total
a. Couldn't attend outside tutoring lectures			2	4	2	1	9
b. Didn't get benefit of doubt						1	1
c. A premedic needs a more general course						1	1
d. Yes, I couldn't work problems					1		1
e. Plenty of chance given in quizzes			2	1	1		4
f. Others higher who deserved less					2		2
g. Last marks should count more					2		2
h. Yes, because I was disgusted and quit					1		1
i. Yes, my quiz marks were low					1		1
j. Difficult for me			2	1	3		6
k. I didn't understand basic laws					1		1
l. I didn't do the work					1		1
m. I didn't know enough to pass					4		4
n. Too many "flunks" in the course				1	1		2
o. Problems in test not fair					1		1
p. Final examination on unimportant material	1	1	1		1		4
q. Difficult and not made interesting			1		1		2
r. No misunderstanding in requirements					1		1
s. I didn't understand and wasn't interested					1		1
t. I didn't understand first part of work				1			1

u. Weekly quizzes not fair			3	3
v. Don't believe in grades			1	1
w. Probably received all I deserved	1[g]	1	4	6
x. My quiz marks were higher			1	1
y. Don't like mathematics and didn't try			1	1
z. A "C" higher than I deserved		2		2
aa. Lowest mark I ever had, but fair on basis used	1	1		2

Discussion of Questionnaire Returns

The distribution of numerical ratings in mechanics for the fall quarter is given in Table LII. The distribution of letter ratings is given in Table LXXVI. These include only those on the original list, those taking one or both pre-tests.

TABLE LXXVI

Distribution of Letter Ratings in Mechanics

Rating	Frequency
A	4
B	18
C	62
D	111
E	43
F	98
I	2
Cancelled	66
No record	20
	424

Questionnaire replies were received from 50 per cent of the A students, 22 per cent of the B students, 23 per cent of the C students, 17 per cent of the D students, 13 per cent of the E, F, and cancelled students, and 25 per cent of those whose records were missing. Those with missing records included those who were taking only laboratory work (and, therefore, took the second pre-test) and a few who were dropped. The seventy from whom replies were received are thus seen to be a fairly representative sample of the whole group. Their opinions are, therefore, probably

representative. A study of the answers given is enough to convince one of their sincerity. It is felt, also, that from the method of sending out the questionnaires desirable frankness is obtained.

In any group of this size there are sure to be some "chronic objectors." These replies are of such a nature, however, as to give the definite impression of a fair-minded, rather unprejudiced group, as a whole. The comments made on whether their marks were fair or unfair are convincing on this point. Forty-five considered their marks fair, while sixteen felt that they were not, and nine had some question in this regard.

As to the feeling of practicality in the course as given, a great majority felt that it could be made more practical. The question, as asked, might be thought suggestive, but the answers given to the next question clarify the matter in this respect.

Question 2 was asked in such a manner as to make no suggestions. The tabulation has been carried out to obviate any possible prejudice on the part of the student. The answers tend to emphasize the matter of impracticability. The greatest number of replies on any one statement—item b—indicates the feeling that the course should be less theoretical, less mathematical, and concern itself more with discussion of the practical. This is emphasized more strongly in items a, c, e, and g. The import of all of these is that the course is so theoretical and mathematical that the members of the class either do not understand clearly, or are not interested because of so much theory and so much lack of stress on the practical aspects.

Undoubtedly, the instructors would have something to say on this question. As one student said (an A student), "I don't think the course is intended to be a very practical one. It is intended more for a theoretical grounding in preparation for the mechanics to be taken later in connection with mathematics in the College of Engineering." The instructors might, or might not, agree with this. In any event a considerable number feel this way about the course, and the fact that there are such low marks and so many failures shows that a great number of students are incapable of doing anything like satisfactory work.

Questions arise here as to the function of the course. Is it designed to select a few capable people from a very large group, who

shall be allowed to go on into engineering, medicine, and the like? Is it to be used as a barrier, or hurdle, to keep the mass of incapables back? Is it intended as a general cultural course to give a broad foundation to a great many? Is it intended to stimulate and inspire a love for and understanding of science and its methods? Is it intended as a specific preparation for engineers? Is it intended as a preparatory course for those who will later teach in high school or college? No attempt will be made here to answer these questions. They are to be answered by those in authority, and student opinion may have some value in arriving at correct solutions. The other answers speak for themselves. Each one tells some sort of a story, which cannot be discussed here in great detail. Good judgments can be distinguished and poor, as well.

Question 3 seeks to get student opinion on the value of laboratory work. The replies seem to be quite in accord with those to the questionnaires given in Chapter II. Items a, b, c, d, e, f, g, and h seem to be similar and outstanding. They emphasize the laboratory as a place to become acquainted with scientific instruments and to gain skill in their use. Items k, l, m, n, o, and p compose another group showing considerable emphasis in that the laboratory is a place where principles should be illustrated, verified, and clarified. The laboratory is thus supplementary to the lectures and tests to accomplish the same purpose but in a more concrete manner. Items r, s, t, u, v, w, x, y, z, and aa emphasize the laboratory as a place to develop certain traits of character useful in life. The interest here is not one of subject matter intrinsically but as a means to the development of character. These three groups are well worth serious thought. Are they mutually exclusive or inclusive? Does the course adapt itself equally well to all three purposes, or should more specific courses be developed, each of which may be designed to accomplish more surely a major objective?

Questions 4 and 5 seek to gain student opinion on the relative values of the various features of science instruction. Replies were somewhat incomplete, but only those that were definite were recorded. Textbook study is seen to be mentioned more times as the most valuable feature, with discussion second, lectures third, and laboratory fourth. The discussions included those given in a tutorial class, held outside the regularly specified hours which only

a part of the students attended, and those held at random by students. From replies to Question 5, only those are listed which indicate the least valuable feature of the instruction. Outside reference reading is considered the least valuable. As a matter of fact, this appears to be a feature which is never prominent. The course is a textual course, not one of much reference reading. The lectures are given as the second least valuable; the quizzes, third.

Question 6 seeks to indicate student interest in physics. The results show much interest, though there are many who would not willingly elect the course.

Question 7 serves to bring out further the bases of interest, or lack of it, on the part of the students. Seventeen merely say that they think it very interesting or they like it. The other replies speak for themselves.

Questions 8 and 9 seek to gain student opinion on the new type of tests used. A majority like the new type better and the outstanding reason is that it covers more ground, thus giving a student more chance to show what he knows. There are, however, a great many preferring the conventional type. The feeling would probably be interpreted to indicate that both types should be used, depending on what purpose they are designed to serve.

Questions 10 and 11 seek to discover student opinion on their marks. The answers show a tendency to quite commendable sportsmanship, which is not confined to A and B students. The replies to these questions, on the whole, give reason for giving serious consideration to the replies to all questions in the questionnaire. They give evidence of sincerity, honesty, and a desire of the students themselves to be fair. Nine mention specifically the outside tutorial class as a feature they feel unfair to them. They seem to feel that these lectures raised the class average rating to such an extent as to lower their individual ratings. This they thought unfair because they were unable to attend these classes on account of the expense involved or the hour of meeting.

CHAPTER XI

THE DEVELOPMENT OF PROGNOSTIC AND ACHIEVEMENT TESTS FOR MECHANICS

The preliminary tests covering the field of high school physics given to all students enrolling in mechanics consisted of portions of partially standardized tests which had been used with high school students for a period of four years in Minneapolis high schools. The highest correlation obtained with these tests was that between the first section of the pre-test and achievement in mechanics when only sophomore engineering students who had taken high school physics were considered. This correlation was $.45 \pm .05$. It was thought that certain carefully selected test items might give ratings which would be much more prognostic than this. Therefore, all preliminary test papers were sorted into three groups; one of students receiving A or B final achievement ratings in mechanics; one of students receiving C ratings in mechanics; and one of students receiving D, E, or F ratings in mechanics. Tables were constructed showing the number and per cent of students in each section giving correct responses to each question in each of the tests. The total number and per cent without reference to sections were also computed.

All questions that did not clearly differentiate the three groups were omitted from the test as finally constructed. For example, Item 1 in Test II, Part I, under the heading "Simple Machines" was omitted because there were 75 per cent of A and B students, 84.48 per cent of C students, and 83.61 per cent of D, E, and F students giving correct responses to it. It was not discriminative. Sixty-three questions were thus omitted and eighty-seven retained. After the retained items were selected according to the method indicated, they were arranged in order of difficulty judged by the per cent of all students giving correct responses. These were then classified as to type, whether short answer, completion, multiple choice, or "X" type, and arranged in test form with the proper direction headings. The test in its final form has items in order of difficulty for each test type, not for the test as a whole. There are

eighty-seven questions in all, constituting information and problem content. The test should prove better for prognosis than the original. Further use must determine the validity of this expectation.

To make use of the three objective examinations given during the fall quarter, 1927, two forms of a final objective examination were constructed. These examinations were given at the end of the quarter's work and may properly be called final examinations. To determine the significance of the questions the papers were separated into three groups, as with the preliminary test: one of A and B students; one of C students; and one of D, E, and F students. These marks—A, B, C, etc., were final ratings obtained at the completion of the course based on all the work of the course. The final letter mark is thus used as a criterion of success in mechanics. The purpose of the plan is to discover and use in final examinations only those questions which have shown differentiating value in selecting A and B; C; and D, E, and F students from the whole group.

Tally sheets were used in finding the number of each group giving correct responses to each question in all three examinations. For convenience these three examinations were called mechanics quiz, Buchta test, and the final objective. The percentage of each group of students getting each question correct was computed, as well as the percentage of the total group. Every question which did not show a decreasing percentage from A students to D, E, and F students, was eliminated from the final forms. Thus, question 3 of the first part of the final objective had 100 per cent of A and B students answering it correctly, 100 per cent of C students, and 89.55 per cent of the D, E, and F students. It was omitted. Question 2 of the same part had 76.19 per cent of A and B students answering it correctly, 35.59 per cent of C students, and 21.82 per cent of D, E, and F students. It was retained. Thirty-six questions were omitted and 115 retained. The types of tests included were short answer, completion, multiple choice, and matching. The questions from all tests of a given type were selected and arranged in order of difficulty as determined by percentages of the total group giving correct responses. The questions were then assigned to Forms A and B, respectively, care being taken not to assign in

regular turn. This was accomplished by taking each group of two questions in turn and alternately assigning the easier first to Form A and second to Form B.

The two tests for prognosis and achievement, together with keys for scoring, have been mimeographed. Information concerning them may be obtained by addressing the Department of Physics, University of Minnesota.

CHAPTER XII

AN EXPERIMENT TO DETERMINE THE EFFECT OF CLASS SIZE IN PHYSICS

During the fall quarter of 1927 an experiment was carried on in the physics course in Heat to determine the effects of size of class on achievement. The classes included 143 students who were originally paired for large and small sections on the bases of average mark points in previous college subjects and scores in a placement test, a modification of the Iowa Placement Test in Physics. For the former criterion, an A was given the value of 1; B, 2; C, 3; D, 4; E, 5; and F, 6. For the latter the raw scores were used. Fifteen students for a small section were paired with fifteen from the large section.

Certain students were unable to arrange their programs to attend the small class and others who could attend were substituted for them. Of fifteen students beginning the small class sessions but nine completed the course with complete records. The original pairing was finally disregarded, necessarily.

Nine is a small number from which to draw conclusions but some value is evident from possible comparisons. As a method of comparison it was thought wise to compare the groups, not only on the original bases of pairing, but on other past records and on their final achievement ratings and final objective tests, as well. It was thought possible, that differences in achievement might be compared with differences in past records and the cause assigned to the one variable—the size of the class.

The past ratings used were as follows: numerical marks in mechanics, scores in the modified Iowa Placement Test in Physics, average mark points, average honor points, and letter marks in chemistry. The difference between mark points and honor points is as follows: Mark points are computed as shown above. Honor points are computed by allowing one point for C, two points for B, and three points for A. D's, E's, and F's are equivalent in this scheme, except that D gives credit in a course, while E and F do not. In the mark-point plan used by Borgersrode[1] in the class-size

[1] F. von Borgersrode, *loc. cit.*

experiments, E's and F's have a definite value in the computations.

In the study of the architectural students in mechanics the marks in chemistry were found to correlate much more highly with achievement in mechanics than any other criterion. Hence, they are used in this study. The letter marks were given the same numerical values as in the computation of mark points. The M scores were not used because all of these ratings were collected before the M score technique occurred to the writer as a desirable possibility. The question arises as to what difference there would be with the weightings, as used, and M scores. To answer this question, M scores in chemistry were computed for these students and the correlation between these M scores and the weighted scores, as explained above, was found to be $.97 \pm .01$. Apparently, for the purpose here, it makes little difference which system of weighting letter marks is used.

For achievement ratings in Heat, the final numerical percentage ratings were used and the ratings in the final objective examination as well. These new type examinations had not been used in this department before and it was thought wise to try them out for use in comparisons, which might be worth while. The final examination consisted of an uncompleted statement type, including a fairly wide sampling from the text used; a true-false test of principles learned in the study of Heat; and a matching biographical test.

The Objective Final Test as a Criterion of Achievement

The regular final examination, which consisted of conventional problem and essay material gave scores which correlated with the objective final test yielding a coefficient of $.38 \pm .05$. This is quite low. The conclusion is that the two tests measure abilities that are related, but which are dependent upon several uncommon factors. For a final mark the mean of these two final examinations was averaged with the mean ratings of ten-hour quizzes given during the quarter, each mean receiving equivalent weight.

The two final examinations—the regular and the objective—were each correlated with the final grade, yielding coefficients of $.76 \pm .03$ and $.63 \pm .04$, respectively. The regular final examination apparently is a more valid criterion of achievement than the objec-

tive final. The multiple correlation[2] with the final grade as the criterion, using the combined scores in the two final examinations, is .84±.02.

The two examinations were more highly correlated together with the final grade than either one alone. It seems wise, then, to continue the use of the final objective examination. There is no apparent reason, however, for using objective examinations exclusively, if the final grades, so obtained, are to be considered the best criteria of achievement. As a matter of fact, the regular type of examination is a more valid measure. These data are of interest in helping us to decide whether or not the new types of examination measure the same functions as the more conventional types. Neither in this study nor in the study of the groups in mechanics do objective examinations prove their superior validity over a five-question problem and essay type. It must be remembered that this applies only to the courses in mechanics and Heat, as taught at the University of Minnesota. The correlations, however, show different factors operating in the two examinations, both groups of factors, however, operating to affect final achievement. Each was considered in calculating the final mark.

In order to find the relationships between the mean rating on the ten-hour quizzes and the two final examinations, the correlation coefficients were computed. The coefficients were .56±.04 for the regular final and .46±.05 for the objective final. The regular final is thus seen to involve more nearly the same functions as the quizzes. A partial correlation may help to show more clearly the real relationship between the objective and regular finals, with the common factors involved in the quizzes eliminated or held constant:

$$r_{12.3} = \frac{.38 - (.46 \times .56)}{(1 - .46^2)\ (1 - .56^2)} = \frac{.122}{.788} = 15.^{3}$$

This would indicate that the factors common to the two types of examination, with the factors involved in the quiz held constant, are represented by a very low coefficient. What is the bearing of this work on experimental work in class size or other studies of the effects of certain variables on achievement? Simply, that the cri-

[2] Otis, *loc. cit.*

[3] *Ibid.*, p. 232.

teria of achievement must be closely scrutinized in the light of objectives of the course, and measurements designed to measure these specific achievements as nearly as possible.

The Comparisons of Groups in Large and Small Classes

Tables LXXVII (a) and LXXVII (b) give the data on group ratings from which comparisons can be made. For these tables the members were paired on the basis of the mark in mechanics, which should be closely related to Heat. The other ratings are given for supplementary comparisons. All students were academic students (College of Science, Literature, and the Arts) and all but one, sophomores. Only one had studied Heat before.

As one studies the data, he is impressed by the evidence that the large group is somewhat superior to the small group. This is evident in all ratings. (Remember that in mark points and ratings in chemistry the larger rating is really lower.) It holds true, also, in achievement in Heat, the greater difference being in the final objective. The inadequacy of pairing data is also evident.

Table LXXVIII is given to show the difference in the means and standard deviations of the whole group of students in the various ratings, for comparison with small and large group sections. The means, standard deviations, and coefficients of variability are to be noted particularly for comparisons with the two groups of nine.

Tables LXXIX (a) and LXXIX (b) show comparisons of the means of all ratings and of individuals. The comparisons of individuals indicate equality of groups in the ratings in mechanics, superiority of the large section in mark points, honor points, chemistry, and the final objective; and superiority of the small section in the placement test and the final grade. The two methods of comparison used should be noted. One uses mean and standard deviation values of groups; the other, the number of members of one group exceeding their paired companions of the equated group. In attempting to generalize on the effects of the small or large classes, the experiment would have to be repeated with many classes, using many different criteria in making the comparisons. This is just what has been done by Borgersrode[4] under the supervision of Hudelson. The results given here do not conflict with their findings in

[4] F. von Borgersrode, *loc. cit.*

TABLE LXXVII (a)

SUMMARY OF THE RATINGS OF STUDENTS IN LARGE SECTION IN HEAT
(Paired Companions of Those in Small Section)

STUDENT	MECHANICS	PLACE-MENT TEST	AVERAGE MARK POINTS	AVERAGE HONOR POINTS	CHEMISTRY	COLLEGE	HEAT BEFORE	FINAL GRADE	FINAL OBJECTIVE
1	61	256	3.34	.88	3.50	A.S.	Yes	64	37
2	72	188	2.76	1.16	3.00	A.S.	No	71	45
3	70	118	2.82	1.29	3.00	A.S.	No	54	22
4	65	201	3.00	.93	3.00	A.S.	No	81	57
5	76	192	2.26	1.76	2.00	A.S.	No	81	41
6	81	174	2.35	1.56	2.00	A.S.	No	78	53
7	59	172	4.02	.98	3.67	A.S.	No	74	49
8	79	223	1.94	2.25	2.00	A.S.	No	84	44
9	80	245	2.23	1.96	2.00	A.S.	No	84	52
Mean	71.44	196.56	2.75	1.42	2.69			74.56	44.44
	2.61**	13.14**	.20**	.15**	.22**			3.18**	3.30**
S. D.	7.82	39.41	.61	.46	.65			9.54	9.89
C. V.	.11	.20	.22	.32	.24			.13	.22

TABLE LXXVII (b)

SUMMARY OF THE RATINGS OF STUDENTS IN SMALL SECTION IN HEAT
(Paired Companions of Those in Large Section)

STUDENT	MECHANICS	PLACE-MENT TEST	AVERAGE MARK POINTS	AVERAGE HONOR POINTS	CHEMISTRY	COLLEGE	HEAT BEFORE	FINAL GRADE	FINAL OBJECTIVE
1	61	162	4.46	.84	4.67	A.S.	No	66	23
2	72	97	4.50	1.00	3.33	A.S.	No	74	33
3	70	133	3.02	1.04	2.50	A.S.	No	59	38
4	65	208	2.58	1.19	3.00	A.S.	No	79	34
5	74	207	2.53	1.37	3.00	A.S.	No	80	50
6	81	187	2.32	1.12	3.00	A.S.	No	82	52
7	60	195	2.88	.89	2.50	A.S.	No	76	39
8	78	190	3.14	1.13	2.00	A.S.	No	75	38
9	81	219	3.43	1.04	4.50	A.S.	No	77	50
Mean	71.33	177.56	3.21	1.07	3.17			74.22	39.67
	±2.52**	±12.56**	±.25**	±.05**	±.28**			±2.30**	±2.99**
S. D.	7.57	37.67	.75	.15	.84			6.89	8.98
C. V	.11	.21	.23	.14	.26			.09	.23

TABLE LXXVIII

SUMMARY OF RATINGS OF ALL STUDENTS ENROLLED IN HEAT AND THOSE OF THE LARGE AND SMALL GROUP SECTIONS

		MECHANICS	PLACE-MENT TEST	AVERAGE MARK POINTS	HONOR POINTS	CHEMISTRY	HEAT	
							Final Obj.	Final Grade
Small group	Mean.....	71.33±2.5	177.56±12.56	3.21±.25	1.07±.05	3.17±.28	39.67±3.0	74.22±2.3
Large group	Mean.....	71.44±2.6	196.56±13.1	2.75±.20	1.42±.15	2.69±.22	44.44±3.3	74.56±3.1
All students	Mean.....	72.77± .88	195.00± 4.9	3.16±.08	1.13±.06	3.04±.11	42.12±1.0	74.71± .89
Small group	S. D.....	7.57	37.67	.75	.15	.84	8.98	6.89
Large group	S. D.....	7.82	39.41	.61	.46	.65	9.89	9.54
All students	S. D.....	9.45	42.00	.70	.47	1.04	11.12	9.78

which they showed that the size of the class apparently does not influence achievement (so far as present measures of achievement are concerned).

From a study of these comparisons, the two samples of nine are seen to be fairly good samples of the whole group. The statement made above—that the sample from the large class is a somewhat superior group—is also confirmed. The superiority extends to all the past ratings as well as to those in Heat. There appears, however, no definite proof that the size of class is a significant factor in achievement.

Relationships among Ratings

For this experiment correlations between achievement and past ratings were not used in order to find the best bases of pairing students. Instead, it was assumed that if many past ratings are used and comparisons made with achievement ratings in Heat, certain differences can be noticed, which may be assigned to size of class. However, the writer has had in mind that one large determining factor, in keeping such correlations low, was the inadequacy of letter ratings in distributing students sufficiently really to show up the correlation actually existing.

Some correlations bearing upon this point have been computed: (1) The scores in the final objective in Heat were reduced to letter ratings, the same percentages of A's, B's, C's, etc., being assigned

TABLE LXXIX (a)

Differences in Means of Paired Groups, Large and Small Sections in Heat

	Mechanics	Placement Test	Average Mark Points	Average Honor Points	Chemistry	Final Grade	Final Objective Test
Difference	.11±3.63*	19.00±18.16*	.46±.29*	.35±.15*	.48±.36*	.34±4.03*	4.77±4.45*
Difference ÷ S.E. of difference	.0	1.0	1.6	2.3	1.3	.1	1.1
Chances in 100 of a difference greater than 0.....	50	84	94	98.9	90	.54	86

* In favor of the large group.

TABLE LXXIX (b)

DIFFERENCES IN MEMBERS OF PAIRED GROUPS, LARGE AND SMALL SECTIONS IN HEAT

(Nine Pairs)

CRITERION	NUMBER AND PERCENTAGE OF STUDENTS* IN SMALL SECTION EXCEEDING PAIRED COMPANIONS IN THE LARGE SECTION		NUMBER AND PERCENTAGE OF STUDENTS IN LARGE SECTION EXCEEDING PAIRED COMPANIONS IN THE SMALL SECTION	
	No.	Per Cent	No.	Per Cent
Mechanics	2	22	2	22
Placement test	5	56	4	44
Average mark points	3	33	6	67
Average honor points	1	11	8	89
Chemistry	4	44	5	55
Final grade in Heat	5	56	4	44
Final objective test	2	22	7	78

* Individual students.

as were given to the same students in mechanics. (See Tables LXXX (a) and LXXX (b).) These letter ratings when correlated with those in mechanics gave a coefficient of .40±.07. This made use of a six-point scale. (2) The numerical score in mechanics was correlated with the final objective in Heat, raw scores, giving a coefficient of .49±.05. This made use of a sixteen-point scale. (3) The numerical final mark in mechanics was correlated with the numerical final mark in Heat, giving a coefficient of .51±.05.

The data seem to lead to the conclusion that the low correlations are not due entirely to insufficient distribution in range but, possibly, to some other factors (as has been stated many times): (1) the lack of really accurate measuring devices; (2) the actual variability of individuals, themselves; (3) the differences due to inherent interest in the particular activity; or (4) the specificity of the activities and the great range of abilities disclosed in an individual due to the great range of specific abilities needed.

The question arises, "Will it ever be possible to get high correlations between achievement in Heat and other criteria?" Seashore[5] has claimed that placement examinations are of more use

[5] Carl E. Seashore, *Learning and Living in College*. University of Iowa Studies, Vol. II, No. 1, chap. iv.

TABLE LXXX (a)

DISTRIBUTION OF MARKS IN MECHANICS GIVEN TO THE MEMBERS OF THE GROUP ENROLLED IN HEAT

RATING	TALLIED DISTRIBUTION	FREQUENCY	PER CENT	APPROXIMATE PER CENT
A	II	2	.9	1
B	IIIIIIIIIIII	13	11.7	12
C	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	40	36.0	37
D	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	39	35.1	35
E	IIIIII	6	5.4	5
F	IIIIIIIIII	11	9.9	10

TABLE LXXX (b)

TRANSMUTATION OF FINAL OBJECTIVE SCORES IN HEAT INTO LETTER MARKS

A 68	C 53—42 inc.	E 26
B 67—54 inc.	D 41—28 inc.	F 25—16 inc.

Based on percentages in Table LXXX (a).

than mental tests for prognosing fitness for, and success in, specific courses. The Iowa Placement Test, given preliminary to the course in mechanics, correlates with the final objective in Heat with a coefficient of but $.32 \pm .08$. This is low and of practically no value for prognosis. It correlates with the final numerical mark in Heat with a coefficient of but $.14 \pm .08$, which is still lower and of less value.

PART IV

CONCLUSIONS

CHAPTER XIII

RECOMMENDATIONS AND CONCLUSIONS

The studies in anatomy, physiology, mechanics, and Heat find their greatest value as actual, concrete illustrations of techniques in educational experimentation. They represent attempts to settle problems of teaching by methods of experiment used in science. They are investigatory in nature, seeking to determine the best methods of attack and the best bases upon which to rest conclusions. No settled technique was used throughout all studies because such a technique has not yet been clearly established. The later studies sought to improve upon the earlier. The studies in anatomy and physiology were made during 1926-27; those in mechanics and Heat during 1927-28. The latter are treated more in detail, partly because of the writer's acquaintance and peculiar interest in them, partly because the studies of 1926-27 had introduced questions in his mind which made more exhaustive treatment desirable. These latter studies may be viewed as introducing procedures, which require much more studied consideration if reliable and valid conclusions are to be made.

For example, the bases of pairing students on ratings in certain psychological tests (mental and reading) and in past records in college were used. The correlations worked out later between these ratings and achievement in the courses of anatomy and physiology, respectively, stressed the question of pairing students. This called forth a more extended investigation and discussion in pairing students in the study in mechanics.

Again, more attention was paid in the later studies to measures of achievement, scaling tests, the effects of high school physics, the differences in students from various colleges, preliminary testing, and student opinion. All of these are intimately connected with studies of this kind. Educational experimentation involves all of these factors, and it seems necessary to include them for consideration rather than to assume, without justification, conditions which eliminate them from consideration. Assumptions have been numerous enough, however, to merit discussion.

In each of the studies there was discernible a certain line of reasoning. In the study in anatomy it was assumed that individuals, who have been paired on the bases mentioned, can be compared in achievement in anatomy and that conclusions can be drawn with reference to differences in the methods of laboratory work. (There is some doubt as to the validity of this assumption, however, and the data were not restricted to this conception.) In addition to treating the individuals as individuals, they were also considered as members of small groups of pairs under a given instructor and as members of large groups of pairs under different instructors who used the same method. Furthermore, all students trained by different methods were compared as groups, whether or not they were paired with other students. The achievement of every student working under one method during the fall quarter was compared with his achievement under the other which was used during the winter quarter. Approximately half of the number worked initially under each method, respectively. These individual records were also compared with individual records of similar groups of students working under the same method for two consecutive quarters of the school year, 1925-26. Moreover, not only the two bases used in the original pairings were used in preliminary comparisons, but other ratings were gathered with which comparisons were made. As to achievement—not merely the final grade but several other bases also were considered.

In the first study in physiology the assumptions were similar, and similar precautions were observed in giving much additional data. In the second study in physiology only paired individuals, as individuals and as groups, were considered, while data are given to show how well these groups are random samples of all the students. In the study in mechanics a serious attempt was made to validate the test ratings and other ratings used, and to find more reliable and valid data for pairing. No preliminary pairing was done. This was delayed until after the completion of the course. Many ratings were given, and these were correlated with achievement in mechanics before being selected as bases for pairing.

No reliable and valid bases of pairing were finally discovered. The comparisons in the best bases found were made, however, by selecting several paired groups to compare with the one group re-

ceiving particular attention, e.g., the architectural group, which had no laboratory work. The number of individuals in the one group exceeding their paired companions in the other compared group was used partially as a criterion of superiority or inferiority. The means and standard deviations of the architectural group were also compared with those of each of the other groups and with the average mean and standard deviations of all the groups.

In the study in Heat only ratings in mechanics were used as a basis of pairing, but several other comparisons were made. The procedure used was that of comparing groups, primarily, in a variety of past ratings and in present achievement. If a striking similarity in groups is evident in many past ratings but not in present achievement, the reason for such dissimilarity should be sought in the known difference in method.

In the studies considered as a whole the following criteria for judgment were used at one time or another.

1. The achievements of individual students paired on one or two bases, which are thought to be best from past experience.

2. The achievements of equated groups based on equality in one or two bases which are thought to be best from past experience.

3. The achievements of equated groups based on many past ratings (including psychological tests), any or all of which may be significant.

4. The achievements of individual students, paired on bases which are proven to be significant.

5. The achievements of groups, equated on bases which are proven to be significant.

6. The achievements of groups, not equated, but which are found comparable on many past ratings and may be compared, therefore, in achievement.

7. The achievements of a group showing a peculiar variation from several other groups which are selected on a similar basis.

8. The achievements of a group judged from that of a smaller group shown to be equivalent to a random sampling.

9. The achievements of individual students exposed in turn to different methods, the effect of one method being obviated by halving the students and using alternative methods with each half initially.

10. The achievements of random samples of students from

classes over a period of years, on the assumption that chance will equate the groups.

No particular criterion was shown to be superior to any other. The highest probability would be expected from the use of as many as possible, each one acting as a check on the other. If any basis of pairing could be proven significant so as to yield a coefficient of .90 or higher, the procedure of pairing students on this basis would be justified. Apparently, there are no such proven bases as yet. There may be in the future, but there is no proof that there will be.

A Suggested Technique

The experience of the writer in these experiments on the college level and in secondary schools leads him to believe that a good, tentative type of experimentation in this field would observe the following essential procedures and precautions:

1. A preliminary period should be given to testing, using training and aptitude tests, general intelligence tests, and other psychological tests. This would include a survey of enrolled students, securing data on age, past training, present environment, present program of activities, and the like.

2. Large enough groups should be employed so that a minimum number of twenty-five per group would have complete records.

3. Selected groups should be on the basis of pure chance.

4. Compare the groups on all available records.

5. Select relatively short periods of experimentation, not to exceed six weeks.

6. Try one method with one group; another method initially with the second group. (There may be more than two groups involved; the more, the better.)

7. At the end of the first half of the experimental period, switch the groups to the alternative method.

8. In instruction keep all factors constant except the one which is being studied.

9. Make comparisons of the individuals under the different methods, studying the two groups in order to discover any effects of the prior method on the subsequent alternative method.

10. Make comparisons of the groups of different students in all data available in regard to past records and present achievement.

11. Exhaust all available means to discover differences and causes therefor.

12. Repeat the experiment, if possible, many times and in different localities with different instructors.

13. Consider as final only conclusions which are repeatedly substantiated.

Conclusions

The more important tentative conclusions from these studies may be summarized as follows:

1. There is no evidence, except individual opinion, to warrant the belief that, under the conditions existing in the Anatomy Department of the University of Minnesota, one cadaver for each two students is more advantageous than one for each four students, as far as the measures of achievement employed could determine.

2. There is some evidence to indicate that seven and one-half hours of laboratory work in human physiology, as carried on in the Medical School of the University of Minnesota, produces measurable achievement in excess of that produced by five hours of laboratory work plus two and one-half hours of library work.

3. It has been shown that the addition of laboratory work in the course in elementary physics (mechanics) at the University of Minnesota on the part of architectural students would raise their achievement records in certain specific items associated with laboratory activities. There is no evidence, however, to show increased achievement in items associated with the other activities of the course.

4. There is additional evidence that class size is not a significant factor in achievement in elementary physics under present instructional conditions. Achievement seems to be more a matter of individual incentive, capacity, and effort.

5. Existing measures of educational products need further study and improvement. Many abilities, presumably developed by science instruction, have not been objectively measured and there is no objective proof that the claims are just.

6. Abilities presumably gained in the laboratory have not been measured objectively as yet, though a beginning was made in this study in the laboratory items incorporated in the second objective test used in mechanics.

7. Subjective opinion largely determines present instructional practices in science.

8. M scores serve as a convenient and valuable means of expressing all ratings. They are then expressed in readily comparable units.

9. Raw scores in classroom tests are as valuable for certain purposes of comparison as scores where individual questions are given a weighted scale value.

10. Evidence shows that the method of weighting letter marks in a given subject has no effect on the correlations obtained when these ratings are used as one variable. No evidence is given, however, to show that the same weighting for all subject letter marks does not affect the correlations.

11. Some objective evidence is given to show that new type tests involve mental functions not involved in the conventional types.

12. A majority of students in the experiment in mechanics preferred objective tests. Judged by the frequency of mention, the primary reason is that they cover more ground and thus give the student more chance to show what he knows. A combination of conventional and objective tests is probably desirable.

13. There is objective evidence in the large percentage of cancellations and failures, that either the subject of mechanics is not adapted to student needs, or, that the students are not prepared to undertake the work.

14. Subjective opinion from three questionnaire returns indicates that the main function of the laboratory is to give the student acquaintance with scientific apparatus and develop skill in its use, and to furnish concrete material for illustration and verification.

15. Student opinion indicates a desire that mechanics may be more practical, more illustrative of applications of physical principles, and less theoretical and mathematical.

16. Evidence shows either that factors other than intelligence are powerful at the college level, or that the intelligence tests do not measure certain features of intelligence which are operative in college work especially.

17. Objective tests in mechanics at the University of Minnesota show that students who have studied high school physics are greatly

superior at the beginning of the course to those who have not studied it.

18. Students enrolled in mechanics who have studied high school physics are not of significantly higher rating in the freshman entrance test scores than those who have not studied high school physics.

19. Achievement in mechanics is significantly higher on the average for those who have had high school physics. The difference is much less, however, than initial differences in the preliminary test covering the field of high school physics. Success in high school physics contributes to success in university physics in this case, but the lack of the former does not prohibit success in the latter. Either large initial differences due to the study of high school physics are much diminished during the course in mechanics, or the content in mechanics is sufficiently different to discount initial differences in high school physics.

PART V

A SUPPLEMENTARY STUDY IN THE TEACHING OF COLLEGE PHYSICS

CHAPTER XIV

SOME POSSIBLE DETERMINANTS OF ACHIEVEMENT IN COLLEGE PHYSICS[1]

Class Size

In the spring quarter of the school year 1927-28 an experiment in class size was carried on in the course of magnetism and electricity offered in the Department of Physics, University of Minnesota. This followed the general plan of several previous experiments in other courses offered in the department; i.e., a small group of students were selected from all those enrolled, and segregated in a small class. The achievements of these students were compared with a similar number enrolled in very large classes to determine if there were any noticeable advantages of the small-class plan over the large-class plan.

In the group enrolled in the course in magnetism and electricity there were three large groups and the one small group. Two of the large groups and the small group were under the immediate direction of Professor Anthony Zeleny while the remaining large group was under the immediate direction of Professor J. W. Buchta. In making comparisons of students in large and small classes, only students from Professor Zeleny's classes were considered so as to obviate any variations due to differences in instructors. All students were included, however, for securing other desirable data to test reliability and validity and obtain M score values.

Procedures for pairing students were similar to those used in the experiment conducted in the class in Heat carried on during the fall quarter of the same year.[2] Students were not paired for comparisons of achievement until the end of the course, and data obtained in regard to the most valuable criteria for pairing. It was assumed from the results of former experiments and correlations, that the most valid data for pairing could be obtained from a study of achievement in relation to other recorded data of the students after the completion of the course. Ratings in mental ability tests

[1] This part was written subsequent to the foregoing chapters.

[2] See chap. xii.

were not used because of low correlations obtained in previous experiments conducted in the department. A preliminary test in elementary electricity and magnetism, devised by the writer and used with classes in high school physics, was given, however, in order to determine if there was any relation between success in the subject in the high school and in the University. This test has a reliability of .92±.01 as computed from the results of its administration to a former class in mechanics in the University and .91±.01 as computed from the results of its administration in this experiment. It correlates, however, but .37±.05 with the final grades in the course in magnetism and electricity, which is somewhat above former correlations obtained from mental ability tests, but not high enough to prognose reliably. The data considered for comparisons for pairing and the correlations obtained are given in Table LXXXI. (M scores are those obtained by using McCall's[3] T score technique for the group of students enrolled.)

TABLE LXXXI

RATINGS USED FOR DETERMINING BEST MEANS OF PAIRING STUDENTS IN LARGE AND SMALL CLASSES IN MAGNETISM AND ELECTRICITY AND CORRELATIONS WITH THE FINAL RATING IN M SCORES

TESTS	SCORES	COEFFICIENT	P.E.	NO. OF CASES
1. Test X	Raw	.37	.05	138
2. Mechanics	M	.49	.04	237
3. Heat	Av. M	.60	.03	162
4. Mechanics and Heat	Av. M	.60	.03	149
5. Mechanics and Heat, Test X	Av. M	.62	.03	143
6. Test X*	M	.38	.05	118

* Only engineers having had high school physics counted.

One would think that achievement in magnetism and electricity could be best predicted from measured achievements in previous courses in college physics or the high school achievement in magnetism and electricity. The fifth correlation in Table LXXXI is the highest, and though not as high as desirable is higher than any obtained in a former experiment in mechanics.[4]

[3] W. A. McCall, *loc. cit.*

[4] *Supra.*, p. 103.

Of fifteen students, in the small class of nineteen who had records suitable to be used in making comparisons, but eight had complete records in all the ratings for which comparisons were especially desirable. The plan of pairing this group of fifteen with an equal number of members of the large classes in all the available ratings was finally adopted. Thus, one student in the small class was paired with a student in a large class in the ratings available for both. Some students were, therefore, paired on the basis of a greater number of ratings. The data given in Table LXXXII (a) will make this clear. Where ratings are missing for a given student, they are missing also for his paired companion. The means and

TABLE LXXXII (a)

COMPARISONS OF LARGE AND SMALL CLASSES IN MAGNETISM AND ELECTRICITY

	Small Class						
Student	Mechanics and Heat, Test X	Mechanics and Heat	Mechanics	Test X	Heat	Mag. & Elec. Final Rating	Mag. and Elec. Objective Final Test
322 SD			66.2	64.5		62.6	59.0
89 BR			62.0			59.0	54.5
97 BH				61.5		49.5	
143 HR			62.0	48.3		52.8	72.0
147 HG			69.5	55.0		62.6	67.7
148 HR	58.7	61.0	62.0	54.0	60.0	65.9	67.7
149 HS	52.2	50.0	45.1	56.5	55.0	59.0	47.5
160 JOH	45.2	49.2	58.0	37.2	40.4	39.1	45.5
183 MG	49.7	47.1	49.2	55.0	45.0	46.3	47.5
185 MW	55.3	54.0	58.0	57.9	50.0	46.3	45.5
193 ME	43.8	44.8	49.2	41.4	40.4	46.3	40.7
199 NR			66.2	50.9		49.5	54.5
224 RE			53.8	44.9		49.5	45.5
230 SR	54.0	56.0	52.0	50.0	50.0	49.5	54.5
248 WS	46.6	54.0	58.0	31.9	50.0	46.3	47.5
Mean	50.69	52.01	58.65	50.64	48.85	52.28	53.54
	±1.74	±1.74	±1.84	±2.10	±2.25	±1.90	±2.29
S. D	4.92	5.06	6.88	8.85	6.36	7.46	8.55
C. V	.097	.097	.120	.175	.130	.143	1.59

TABLE LXXXII (a)—*Continued*

COMPARISONS OF LARGE AND SMALL CLASSES IN MAGNETISM AND ELECTRICITY

	Large Class						
Student	Mechanics and Heat, Test X	Mechanics and Heat	Mechanics	Test X	Heat	Mag. & Elec. Final Rating	Mag. and Elec. Objective Final Test
155 JJF			66.2	59.0		55.9	63.3
236 SSR			62.0			49.5	40.7
205 PAD				59.0		59.0	
133 FB			62.0	50.0		46.3	40.7
321 SR			69.5	50.0		70.0	63.3
135 GHD	58.2	56.5	58.0	61.5	55.0	65.9	63.3
272 EA	51.3	48.6	62.0	56.5	35.3	49.5	54.5
254 WJ	46.0	51.2	62.0	35.7	40.4	46.3	45.5
227 SJW	49.8	53.5	62.0	41.4	45.0	49.5	47.5
161 JTR	55.7	54.0	58.0	59.0	50.0	46.3	45.5
175 LGE	43.8	47.1	49.2	37.2	45.0	49.5	45.5
92 BL			66.2	39.1		52.8	54.5
103 BVR			53.8	41.4		39.1	40.7
146 HCF	54.8	54.0	58.0	56.5	50.0	59.0	54.5
142 HEB	46.9	49.6	49.2	41.4	50.0	46.3	30.8
Mean	50.81	51.81	59.86	49.12	46.33	52.32	49.30
	±1.62	±1.06	±1.55	±2.27	±2.08	±2.06	±2.43
S. D	4.79	3.00	5.79	8.49	5.88	7.96	9.09
C. V	.094	.058	.097	.173	.127	.152	.184

standard deviations are not always computed for the whole number of fifteen for all ratings but are always computed for the same number in the small class as in the large. All the ratings are expressed in M scores, 50 always representing the mean of the whole group and 10 the S.D.

Two means of comparison are possible: (1) a comparison of whole groups; and (2) a comparison of individuals within the groups. The first will be accomplished by comparing the means and coefficients of variabilities of the two groups. Table LXXXII (b) is given to make comparisons of the means.

The differences are insignificant statistically with the exception of that between the means of Test X scores. Here the small class

TABLE LXXXII (b)

COMPARISONS OF THE MEANS OF THE SMALL AND LARGE CLASS RATINGS IN THE CRITERIA OF TABLE LXXXI

	DIFFERENCE IN MEANS	S. E. OF DIFFERENCE	DIFFERENCE ÷ S. E. OF DIFFERENCE
Mechanics, Heat, and Test X	.12	2.38	.0
Mechanics and Heat	.20	2.04	.1
Mechanics	1.21	2.44	.5
Text X	1.52	3.09	.5
Heat	2.52	3.06	.8
Magnetism and Electricity, Final Rating	.04	2.74	.0
Magnetism and Electricity, Objective Test	4.24	3.34	1.3

showed a higher mean, but it will be remembered from Table LXXXI that this was the least valuable criterion used in pairing students. The small class also had a higher mean in the objective test given near the end of the quarter. This test was quite comprehensive, consisting of seventy-seven items with a possible score of 79. The superiority was not significant statistically, however, so we must conclude that no significant differences show up between the large and small class groups. Especially in the final marks the means of the two groups were practically equal.

Data for judging individuals are found in Table LXXXIII.

TABLE LXXXIII

COMPARISONS OF ACHIEVEMENT RATINGS OF LARGE AND SMALL GROUPS GIVEN IN THE CRITERIA OF TABLES LXXXII (a) AND LXXXII (b)

	NUMBER IN LARGE SECTION EXCEEDING PAIRED COMPANIONS IN SMALL SECTION	NUMBER IN SMALL SECTION EXCEEDING PAIRED COMPANIONS IN LARGE SECTION
Test X, Mechanics and Heat	5	2
Mechanics and Heat	4	4
Mechanics	3	3
Test X	5	7
Heat	1	2
Magnetism and Electricity (final mark)	7	5
Magnetism and Electricity (objective test)	3	6

An examination of Table LXXXIII brings out no striking differences which may be assigned to the effects of large or small classes. From the numbers exceeding their paired companions in the first criterion, i.e., ratings in mechanics, Heat and Test X, combined, one might conclude that the members of the large group were markedly superior to those of the small. However, the data in Tables LXXXII (a), LXXXII (b), and LXXXII (c) show that the differences are practically nothing, as the members of the two groups were primarily paired on this particular basis. In criteria 4 and 5 the small group seems somewhat superior. In the final ratings in magnetism and electricity the large group appears slightly superior though in the case of one student the superiority is very small. In the objective test ratings, however, the small group shows a more marked superiority, which tends to even up the apparent superiority of the large group in the final rating.

The question of which rating is a more valid criterion of success in the course naturally arises. The proponents of objective testing would say that the objective-test ratings are more valid, assuming that the objective test is carefully devised. There seems to be no way of determining the answer to the question of validity. If it be assumed, as seems necessary, that the final mark in the course is the most valid criterion of success, the objective test is not so valid, because its correlation with the final mark is but $.66 \pm .02$. The final conventional test, which has not been included separately in the tables so far given, correlates much higher with the final mark, the coefficient being $.87 \pm .01$. In other words, one conventional examination, viz., the final examination, correlates more highly with the final mark than the objective examination. Therefore, its measured validity is higher. The proponents of objective tests as measures of achievement would probably say, however, that the assumption that the final mark is the best criterion of success is an erroneous assumption. The final marks in this case were computed from scores on conventional tests given once a week, together with the scores on the final conventional examination and scores on the one objective test. The fact that the final conventional correlates so highly with the final mark is not surprising, under the circumstances. It indicates that the conventional test may be quite objective, i.e., the scoring may be done in such a manner that it

becomes objective, or, that the quiz scores for the quarter are quite consistent and reliable. It must be kept in mind that scores may be reliable, however, without being valid, i.e., the conventional tests might be reliable in placing a student at a given place in a group without being valid in measuring the most desirable outcomes of the course. When objective tests and conventional tests do not agree, it is well to make a careful study of each and decide after careful analysis which test measures the most desirable aspects. The coefficient of reliability for the objective test, computed by considering the even-numbered and odd-numbered items as two tests covering the same subject, finding the scores in each set, and correlating them is .73±.02. Correcting the coefficient by the Spearman-Brown formula,[5] gives a coefficient of .85±.01. This is not as high as it should be to call the test highly reliable. The test may need refinement to eliminate some possibly ambiguous statements.

The question of reliability of subjective tests is important. If the conventional final examination correlates more highly with the final rating than does the objective final examination, its validity is higher, using the final rating as the criterion. Does the conventional examination have higher reliability also? Its reliability cannot be determined by correlating scores on the even- and odd-numbered questions because there are too few questions. A judgment of the reliability of the conventional quizzes given during the quarter, however, may be obtained from a correlation between the average scores of Tests I, III, V, and VII and Tests II, IV, VI, and VIII. This yields a coefficient of .68±.03. This is the reliability coefficient for a three-hour conventional examination of the type given in this course. The final examination is but a two-hour examination. Assuming its reliability as high as the quizzes and its length two-thirds as long, its reliability would be less than .68, which is much lower than the single objective examination.

It might be worth while to show the distributions for the conventional and objective tests, respectively. The form of distribution of scores is indicative of some characteristics of a test. A tallied distribution gives some idea of this form without entering into the statistical procedures necessary to determine how closely

[5] H. E. Garrett, *Statistics in Psychology and Education.*

a distribution resembles a normal distribution. Tables LXXXIV and LXXXV give these distributions together with M values. An examination of the distributions makes the conventional, final examination appear fairly good, as far as the distribution of scores is concerned. There is no evidence showing the objective examination superior in this respect.

TABLE LXXXIV

DISTRIBUTION OF FINAL CONVENTIONAL TEST SCORES AND M VALUES IN MAGNETISM AND ELECTRICITY

SCORES	TALLIED DISTRIBUTION	FREQUENCY	M VALUES
0– 4.9	III	3	24.2
5– 9.9	I	1	27.3
10– 14.9	I	1	28.2
15– 19.9	II	2	29.4
20– 24.9	II	2	30.6
25– 29.9	IIIIIIIIIII	11	33.3
30– 34.9	IIIIIIIIIII	11	36.2
35– 39.9	IIIIIIIIIIIIIIII	16	38.6
40– 44.9	IIIIIIIIIIIIIIIIIIIIIIIIII	26	41.5
45– 49.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIII	29	44.3
50– 54.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	31	46.9
55– 59.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	32	49.6
60– 64.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIII	30	52.2
65– 69.9	IIIIIIIIIIIIIIIIIIII	22	54.5
70– 74.9	IIIIIIIIIIIIIIIIIIIIIIIII	27	56.8
75– 79.9	IIIIIIIIIIIIIIIIIIII	22	59.5
80– 84.9	IIIIIIIIIIIIIIII	17	62.5
85– 89.9	IIIIIIIII	10	65.4
90– 94.9	IIIIIIII	9	68.6
95– 99.9	IIII	4	73.4
100–104.9	I	1	79.5
	Total	307	

For the benefit of those who may feel that other means of equating students should have been investigated in this study, the M scores in chemistry, mathematics, and English were computed for the students enrolled in magnetism and electricity who had taken these subjects. The percentile ratings in the freshman mental

TABLE LXXXV

DISTRIBUTION OF SCORES IN FINAL OBJECTIVE EXAMINATION IN MAGNETISM AND ELECTRICITY

SCORES	TALLIED DISTRIBUTION	FREQUENCY	M VALUES
20–24.9	IIII	4	26.0
25–29.9	IIIIII	6	30.8
30–34.9	IIIIIIIIIIIIIIIIII	19	35.8
35–39.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIII	31	40.7
40–44.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	43	45.5
45–49.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	46	47.5
50–54.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	42	54.5
55–59.9	IIIIIIIIIIIIIIIIIIIIIIIII	27	59.0
60–64.9	IIIIIIIIIIIIIIII	18	63.3
65–69.9	IIIIIIIII	9	67.7
70–74.9	III	3	72.0
75–79.9	II	2	76.5
	Total	250	

TABLE LXXXVI

COEFFICIENTS OF CORRELATION WITH FINAL RATINGS IN MAGNETISM AND ELECTRICITY WHICH MIGHT HAVE BEEN CONSIDERED IN PAIRING STUDENTS FOR THIS EXPERIMENT

RATING	COEFFICIENT	P.E.	NO. OF CASES
Freshman test percentiles	.26	.06	128
Chemistry M scores	.44	.05	122
Mathematics M scores	.41	.05	133
English M scores	.16	.06	120

tests, given to all entering students in the College of Science, Literature, and the Arts, were also obtained. Correlations were computed between the final achievement ratings in the course in magnetism and electricity, and each of the other ratings mentioned. Table LXXXVI gives these correlations. None of these coefficients are as high as those actually used, with the exception of the preliminary test in high school content in magnetism and electricity. It will be remembered that this was not the major criterion used, however.

Factors of Previous Courses in the Secondary School

Always an interesting phase of a study of achievement in the University is the relationship of this achievement to previous work in the secondary school. For example, do students who have studied physics in the high school do better work in college classes in physics? Data on this matter have been gathered and Table LXXXVII presents them. There is seen to be a difference in the

TABLE LXXXVII

Distribution of Ratings in Magnetism and Electricity

Students Having Had H. S. Physics		Total	Students Not Having Had H. S. Physics		Total
35– 37.9	III	3	35– 37.9		0
38– 40.9	I	1	38– 40.9	I	1
41– 43.9	I	1	41– 43.9		0
44– 46.9		0	44– 46.9	I	1
47– 49.9	I	1	47– 49.9		0
50– 52.9	IIIIIIIII	9	50– 52.9	II	2
53– 55.9	IIIIIIIII	9	53– 55.9	II	2
56– 58.9	IIIIIII	7	56– 58.9	III	3
59– 61.9	IIIIIIIIIIIIIIIIIIII	20	59– 61.9	IIIIIIIII	9
62– 64.9	IIIIIIIIIIIIIIIII	18	62– 64.9	IIIIII	6
65– 67.9	IIIIIIIIIIIIII	15	65– 67.9	III	3
68– 70.9	IIIIIIIIIIIIIIIIIIII	22	68– 70.9	IIIIII	6
71– 73.9	IIIIIIIIIIIIII	16	71– 73.9	IIIIIII	7
74– 76.9	IIIIIIIIIII	12	74– 76.9	II	2
77– 79.9	IIIIIIIIIIII	13	77– 79.9	III	3
80– 82.9	IIIIIIIIIIIIIIII	18	80– 82.9	III	3
83– 85.9	IIII	4	83– 85.9	I	1
86– 88.9	IIIIIIIII	9	86– 88.9		0
89– 91.9	IIIIII	6	89– 91.9		0
92– 94.9	III	3	92– 94.9		0
95– 97.9	I	1	95– 97.9		0
98–100.9	I	1	98–100.9		0
		189			49

Mean	69.70±.87	Mean	66.19±1.35
S. D.	12.09	S. D.	9.49
C. V.	.17	C. V.	.14

Difference in means, 3.51 ± 1.61

Diff. ÷ S.D. of diff. = 2.2

means, which may not be considered significant by some authorities, but which is highly probable. It may be stated that, on the average, students who have had high school physics will probably do better work in magnetism and electricity in the University of Minnesota, although not all students who have had the high school course do better than those who have not had the course. Sixteen out of forty-nine, who have not had high school physics exceed the mean (final rating in magnetism and electricity) of those who have had high school physics.

If students who have studied high school physics do not show a superiority in college physics, it may be reasonable to ask what values are obtained from high school physics. Test X, a test covering the field of magnetism and electricity in high school physics, was administered to this university class, as stated in a previous paragraph, and the scores in this test are indicative of a knowledge of the content covered in high school. Table LXXXVIII gives a distribution of scores in this test earned by the students enrolled in magnetism and electricity during the spring quarter of 1928. The norm (median) obtained from the administration of this test

TABLE LXXXVIII

DISTRIBUTION OF SCORES IN HURD TEST X OF STUDENTS ENROLLED IN MAGNETISM AND ELECTRICITY

SCORES	TALLIED DISTRIBUTION	FREQUENCY
2– 4.9	IIIIII	6
5– 7.9	IIIIIIIIIIIIIIIIIIIII	23
8 10.9	IIIIIIIIIIIIIIIIIIIIIII	25
11–13.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	43
14–16.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	45
17–19.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	39
20–22.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	44
23–25.9	IIIIIIIIIIIIIIIIIIIIIIII	27
26–28.9	IIIIIIIIIIIIIIIIIIIIIIII	27
29–31.9	IIIIIIIIIIIIIIIIII	20
32–34.9	IIIIIIIIIIIIIIII	17
35–37.9	IIIIIII	7
		323

Mean = 19.04
S. D. = 8.07

to 350 high school students is 20.09. Apparently, the students in the college class at the beginning of the course are, on the average, about on the level of high school students in the content covered by Test X.

Table LXXXIX gives distributions of ratings in Test X of students who had taken a high school course in physics, and those who had not taken a high school course. There is a great difference between those having had and those not having had high school physics. This is statistical evidence showing some value at least for the average student in studying physics in high school. Here again it is noticed that twelve students out of seventy-three of those not having had high school physics exceed the mean of those having had it. Apparently, in general the average student learns some facts from high school physics which stay with him for an indefinite period of time, which he who has not had the high school course does not learn. However, this average student who has had the high school course does not show very great superiority in the

TABLE LXXXIX

Distribution of Scores in Hurd Test X in Magnetism and Electricity

Students Having Had H. S. Physics		Total	Students Not Having Had H. S. Physics		Total
2– 4.9	I	1	2– 4.9	IIIIII	6
5– 7.9	IIIIIIII	8	5– 7.9	IIIIIIIIIIIII	14
8–10.9	IIIIIIIIIIIIIIIII	17	8–10.9	IIIIIIIIII	10
11–13.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	32	11–13.9	IIIIIIIIII	11
14–16.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	40	14–16.9	IIIIIIIIII	10
17–19.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	33	17–19.9	IIIIIIIIII	10
20–22.9	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	37	20–22.9	IIIIII	6
23–25.9	IIIIIIIIIIIIIII	17	23–25.9	II	2
26–28.9	IIIIIIIIIIIIIIIIIIIIIII	24	26–28.9	II	2
29–31.9	IIIIIIIIIIIIIIII	18	29–31.9	I	1
32–34.9	IIIIIIIIIIIIII	16	32–34.9	I	1
35–37.9	IIIIIII	7	35–37.9		
		250			73
Mean = 20.24 ± .49			Mean = 13.44 ± .79		
S. D. = 7.71			S. D. = 6.87		
Difference in means, 6.80 ± .93					
Diff. ÷ S. E. of diff. = 7.3					

content of the university course in magnetism and electricity at Minnesota at the completion of this course. Two possible reasons may be assigned for this. First, the content of the university course may be so unlike the high school course that no benefits accrue to the student who has taken the high school course. For example, the difficulties and complexities of the university course may be those of higher mathematics. Second, the student who enters the university course without high school physics, may be able to overcome nearly all of the deficiencies he may have by the time the course is completed.

TABLE XC

ITEMS IN PRELIMINARY TEST X ANSWERED CORRECTLY BY 90 PER CENT OF STUDENTS IN MAGNETISM AND ELECTRICITY

	No.	PER CENT
1. How many poles has a compass needle?	317	95.0
3. From the following select a non-magnetic substance: copper, silver, steel, gold, zinc, lead	328	98.7
5. Name the international unit of electric resistance	304	91.5

What are the items to which correct responses were given by the students enrolling for the university course in magnetism and electricity? But three of those included in Test X were answered correctly by 90 per cent or more of these students. They are given in Table XC together with the number and percentage of the students giving correct responses. It seems almost absurd to think that of 323 students taking this test including 250 who had studied high school physics, but three simple questions of 42 given were answered correctly by 90 per cent or more of such a group. Twenty-six of the forty-two questions were failed by 50 per cent or more of the group. Table XCI gives these items.

Table XCII contains those items which were failed by 75 per cent or more of these students. If only those students who had studied high school physics are considered, the number of items to which correct responses were given would undoubtedly be larger. This is indicated by the difference in means of those having had, and those not having had high school physics, as shown in Table LXXXIX.

TABLE XCI

ITEMS IN PRELIMINARY TEST X FAILED BY AT LEAST 50 PER CENT OF STUDENTS IN MAGNETISM AND ELECTRICITY

		No.*	PER CENT†
2.	If a current of .5 amperes is passing through each of two 110 volt incandescent lamps in series, find the total impressed voltage	133	40.0
3.	Find the resistance of the same two lamps in parallel	99	29.8
4.	If a transformer whose efficiency is 100% has 11,000 turns of wire on its primary coil receiving an E.M.F of 2,200 volts, how many turns of wire are theoretically necessary on its secondary coil, so that the E.M.F. of the output shall be 110 volts?	121	36.4
5.	A carbon incandescent 110 volt lamp carries a current of ½ ampere; what it is power consumption?	138	41.5
6.	Find the current in two 110 volt, 220 ohm lamps in series	62	18.7
7.	If a current of .5 ampere is passing through each of two lamps in parallel, find the total current	75	22.6
8.	Find the efficiency of a transformer whose input is 5 amperes at 2,200 volts and whose output is 100 amperes at 100 volts	103	31.0
9.	How much resistance is needed in series with an arc lamp having a resistance of 5 ohms on a service line of 110 volts E.M.F. to produce a current of 10 amperes?	89	27.8
10.	Find the total impressed voltage for two 220 ohm incandescent lamps in parallel if each carries .5 ampere	55	16.5
12.	Give the names of one of the magnetic poles	11	3.3
14.	Give the names of the static charges of electricity	108	32.5
15.	Using the *names* of the *units,* give the formula representing Ohm's Law	68	20.4
16.	What is the name of the process by means of which an iron rod is magnetized by being held in the position of a dip needle and hammered?	28	8.4
17.	According to the theory of magnetism, what are the molecules of all magnetic substances?	70	21.0
18.	What magnetic pole of the earth is found in the vicinity of Hudson Bay?	30	9.0
19.	How large a resistance should an ammeter have?	83	25.0
23.	What material constitutes the anode in a nickel-plating outfit?	155	46.7

* Number giving correct responses.

† Percentage giving correct responses.

TABLE XCI—*Continued*

ITEMS IN PRELIMINARY TEST X FAILED BY AT LEAST 50 PER CENT OF STUDENTS IN MAGNETISM AND ELECTRICITY

		No.*	PER CENT†
24.	Give the formula for finding the resistance of a wire of known material, length, and diameter	2	.6
25.	Name the part of a dynamo in which the induced current is produced	118	35.5
26.	What kind of energy is obtained from an electric motor?	145	43.6
27.	If a current through a north and south conductor above a magnetic needle turns the N. Pole toward the east, what is the direction of the current?	82	23.7
28.	What essential part of a direct current generator is not a part of a simple, alternating current generator?	65	19.6
29.	What device is commonly used to increase the voltage of direct currents?	17	5.1
30.	What device is commonly used to change an alternating current to a direct current?	78	23.5
31.	How is the efficiency of an incandescent lamp usually expressed?	15	4.5
32.	From the following select the best conductor of electricity: iron, copper, zinc, silver, gold, mercury	123	37.0

* Number giving correct responses.

† Percentage giving correct responses.

Table XCIII gives the items to which there were 90 per cent or more correct responses on the part of students enrolled in the course in magnetism and electricity who had taken high school physics. While there were but three items answered correctly by 90 per cent or more of the students when all were counted (see Table XC), two more are added when all those not having had high school physics are excluded.

A list of items failed by 50 per cent or more of those students who had taken high school physics was prepared but it is almost identical with Table XCI, there being but one less item (26 and 25, respectively). Here, there seems practically no difference between those having had and those not having had high school physics. A list was also prepared showing items failed by 75 per cent or more of the students having had high school physics. This

TABLE XCII

ITEMS IN PRELIMINARY TEST X FAILED BY AT LEAST 75 PER CENT OF STUDENTS IN MAGNETISM AND ELECTRICITY

		No.*	PER CENT†
6.	Find the current in two 110 volt, 220 ohm lamps in series	62	18.7
7.	If a current of .5 amperes is passing through each of two lamps in parallel, find the total current	75	22.6
10.	Find the total impressed voltage for two 220 ohm incandescent lamps in parallel if each carries .5 ampere	55	16.5
12.	Give the names of one of the magnetic poles	11	3.3
15.	Using the *names* of the *units*, give the formula representing Ohm's Law	68	20.4
16.	What is the name of the process by means of which an iron rod is magnetized by being held in the position of a dip needle and hammered?	28	8.4
17.	According to the theory of magnetism, what are the molecules of all magnetic substances?	70	21.0
18.	What magnetic pole of the earth is found in the vicinity of Hudson Bay?	30	9.0
19.	How large a resistance should an ammeter have?	83	25.0
24.	Give the formula for finding the resistance of a wire of known material, length, and diameter	2	.6
28.	What essential part of a simple direct current generator is not a part of a simple alternating current generator?	65	19.6
29.	What device is commonly used to increase the voltage of direct currents?	17	5.1
30.	What device is commonly used to change an alternating current to a direct current?	78	23.5
31.	How is the efficiency of an incandescent lamp usually expressed?	15	4.5

* Number giving correct responses.

† Percentage giving correct responses.

shows ten items as contrasted with fourteen when all students are counted, thus indicating some value for high school physics.

A comparison of these lists of items that were prepared, for students not having had high school physics, shows but one item included in the list having 90 per cent or more correct responses; 33 items included among those failed by 50 per cent or more; and 26 items among those failed by 75 per cent or more. These show

TABLE XCIII

ITEMS IN PRELIMINARY TEST X ANSWERED CORRECTLY BY 90 PER CENT OF STUDENTS IN MAGNETISM AND ELECTRICITY HAVING HAD HIGH SCHOOL PHYSICS

	No.*	PER CENT†
1. How many poles has a compass needle?	255	97.1
2. What do like magnetic poles when placed near each other do?	248	94.5
3. From the following select a non-magnetic substance: copper, silver, steel, gold, zinc, lead	261	99.4
4. From (3) select a magnetic substance	239	91.1
5. Name the international unit of electric resistance	251	95.6

* Number giving correct responses.

† Percentage giving correct responses.

considerable differences between groups having had and not having had high school physics. They show also a higher percentage of students who have had high school physics giving correct responses. Table XCIV is given to show differences in percentages of (1) all students, (2) students having had high school physics, and (3) those not having had high school physics, on all items in Test X.

An examination of this table (XCIV) shows that on every test item but one a greater percentage of those having had high school physics gave correct responses than of those not having had high school physics. The one exception is item 24. There were but two correct responses given to this item, one of them being from a student not having had high school physics. The exception is thus accounted for. The means of the percentages of correct responses is seen to be much higher for those having had high school physics than for those not having had it. The coefficient of variation of the percentages for those not having had high school physics is half again as much as that for those having had high school physics. High school instruction in physics, then, produces a greater number of correct responses to items in Test X on the average, and there is less variability in the percentages on the various items, i.e., more homogeneity in these percentages. The constant factors involved in instruction thus apparently have a noticeable effect in raising achievement and lessening the variability

of responses. A pertinent inquiry would be one to determine how many more items and how many more correct responses per item would be listed if Test X had been repeated at the end of the university course in electricity and magnetism and similar lists made

TABLE XCIV

COMPARISONS OF PERCENTAGES OF STUDENTS GIVING CORRECT RESPONSES TO ITEMS IN PRELIMINARY TEST X

TEST ITEM	ALL STUDENTS	STUDENTS HAVING HAD H. S. PHYSICS	STUDENTS NOT HAVING HAD H. S. PHYSICS
Part I			
1	95.0	97.1	87.4
2	89.4	94.5	73.3
3	98.7	99.4	94.5
4	89.7	91.1	83.2
5	91.5	95.6	74.7
6	68.3	73.5	47.9
7	68.0	76.6	36.7
8	64.4	73.5	29.6
9	72.8	83.1	33.8
10	51.5	58.3	25.4
11	79.2	83.4	62.0
12	3.3	3.4	2.8
13	68.9	70.9	60.6
14	32.5	38.5	9.9
15	20.4	24.8	4.2
16	8.4	9.9	2.8
17	21.0	25.5	4.2
18	9.0	9.9	5.6
19	25.0	28.6	11.3
20	70.1	73.5	56.4
21	80.4	83.1	69.1
22	50.9	58.7	21.1
23	46.7	53.7	19.7
24	.6	.4	1.4
25	35.5	38.5	24.0
26	43.6	49.5	21.1
27	25.7	27.1	15.5
28	19.6	22.0	7.0
29	5.1	4.6	7.0
30	23.5	27.4	8.5
31	4.5	5.3	1.4
32	37.0	41.9	18.3

TABLE XCIV—*Continued*

COMPARISONS OF PERCENTAGES OF STUDENTS GIVING CORRECT RESPONSES TO ITEMS IN PRELIMINARY TEST X

TEST ITEM	ALL STUDENTS	STUDENTS HAVING HAD H. S. PHYSICS	STUDENTS NOT HAVING HAD H. S. PHYSICS
Part II			
1	76.2	77.7	69.1
2	40.0	43.4	26.8
3	29.8	32.4	19.7
4	36.4	41.9	15.5
5	41.5	48.8	14.1
6	18.7	21.3	8.5
7	22.6	25.9	9.9
8	31.0	37.3	7.0
9	27.8	31.6	8.5
10	16.5	20.6	1.4
Mean	44.4	47.7	28.6
S. D.	28.3	29.9	26.6
C. V.	.64	.63	.92

as those already referred to. Test X was not given then, however.

Whenever lists are made showing the percentage of students who give correct responses to test items or numbers of test items failed by certain percentages of students, the great discrepancies between objectives and achievements are evident. These discrepancies are jarring to conscientious instructors and raise many questions. Are our objectives too high; or are achievements too low? Have we a legitimate reason for expecting higher achievements? Are we selecting certain legitimate objectives and compelling certain desired degrees of performance? Are we expecting any high degree of achievement? Is education an opportunity or a requirement? Are teachers guides or taskmasters? Certainly if requirements are set up as represented in test items, of which a high degree of achievement is expected, such surveys as these must bring great disillusionment. To be sure, the test items so far considered have been those included in the preliminary test covering the field of high school physics. But are there not similar conditions at university levels?

To answer this question, a tabulation of items in the final objec-

TABLE XCV

ITEMS IN THE FINAL OBJECTIVE EXAMINATION IN MAGNETISM AND ELECTRICITY, WITH NUMBER AND PER CENT OF STUDENTS MAKING CORRECT RESPONSES TO EACH ITEM

	No.	Per Cent		No.	Per Cent		No.	Per Cent		No.	Per Cent
1	202	67.3	22	95	31.6	38(1)	275	91.6	42	145	48.3
2	123	41.0	23	217	72.3	(2)	286	95.2	43	65	21.6
3	241	80.3	24	179	59.6	(3)	236	78.6	44	245	81.6
4	239	79.6	25	113	37.6	(4)	217	72.3	45	265	88.2
5	256	85.2	26	196	65.3	(5)	274	91.2	46	206	68.6
6	193	64.3	27(1)	21	7.0	(6)	263	87.6	47	149	49.6
7	147	49.0	(2)	28	7.3	(7)	260	86.6	48	146	48.6
8	208	69.3	(3)	18	6.0	(8)	225	74.9	49	178	59.3
9	251	83.6	(4)	6	2.0	(9)	186	61.9	50	129	43.0
10	227	75.6	28(1)	35	11.6	(10)	248	82.6	51	204	67.9
11	119	39.6	(2)	87	29.0	(11)	250	83.2	52	182	60.6
12	238	79.3	(3)	25	8.3	(12)	206	68.6	53	43	14.3
13	146	48.6	29	62	20.6	(13)	220	72.3	54	85	28.3
14	89	29.6	30	16	5.3	(14)	253	84.2	55	207	68.9
15	164	54.6	31	61	20.3	(15)	0		56	5	1.7
16	63	21.0	32	196	65.3	(16)	207	68.9	57	156	51.9
17	78	26.0	33	39	13.0	(17)	253	84.2	58	43	14.3
18	138	46.0	34	145	48.3	(18)	250	83.2	59	105	35.0
19	246	81.9	35	83	27.6	39	84	28.3	60	191	63.6
20	255	84.5	36	116	38.6	40	298	82.6			
21	168	55.9	37	57	19.0	41	66	22.0			

tive test in electricity and magnetism, given in the spring of 1928, was carried out. Table XCV gives these data.

An examination of the table indicates no more rosy picture of achievement than is indicated by the high school test items. On the test of seventy-seven items, achievement leaves much to be desired, if these items represent goals of desired achievement. It seemed worth while to prepare a list of (1) items having 90 per cent or more of correct achievement, (2) items failed by 50 per cent or more, (3) items failed by 75 per cent or more, together with the number and per cent giving correct responses to each. This was done but the table is not given here for lack of space. There were shown to be, however, but three items in the list having 90 per cent or more correct responses; thirty-four items failed by 50 per cent or more; and fifteen items failed by 75 per cent or more. These items are on file in the Bureau of Educational Research.

Factor of Interest in Science

A factor which might be thought to have a determining effect on achievement in university physics is the number of sciences studied in high school. This might be thought to be so for at least two reasons: (1) The values directly obtained from the high school courses might carry over into university physics. (2) The number of sciences might be an index of interest in science which might be productive of superior achievement in university physics. In order to secure data on this phase, records of those sciences studied in high school were obtained, and subsequent ratings in university physics tabulated for those students having had no high school science, one high school science, two high school sciences, etc. Table XCVI gives the evidence for Test X (the preliminary test used in this experiment), the final achievement ratings in the university course in magnetism and electricity, and ratings in the university courses in mechanics and Heat, as well. While in Test X and Heat, there seems to be a slight tendency for higher scores when more sciences are studied, there is no conclusive evidence of noteworthy differences. Sciences do not show common abilities which carry over into future courses in other sciences; nor does the number of sciences taken reliably indicate interest which produces greater achievement in future courses. Interest may produce greater

TABLE XCVI

MEANS AND STANDARD DEVIATIONS OF CERTAIN RATINGS IN PHYSICS FOR STUDENTS HAVING TAKEN VARIOUS NUMBERS OF HIGH SCHOOL SCIENCE COURSES

RATINGS		NUMBER OF SCIENCES IN HIGH SCHOOL						ALL STUDENTS
		One	Two	Three	Four	More Than Four	None	
Test X	Mean	16.44	19.47	19.20	21.46	24.16	14.75	19.04
	S. D.	9.06	7.51	7.89	8.42	6.62	8.10	8.07
	Cases	46	142	77	29	11	16	323
Heat	Mean	80.75	78.29	79.29	81.37	82.00	79.84	79.20
	S. D.	10.35	8.91	6.90	6.78	9.99	6.69	8.4
	Cases	12	82	42	16	9	9	166
Mechanics	Mean	68.40	69.42	71.26	71.09	69.20	67.75	69.51
	S. D.	10.23	9.45	9.21	9.63	9.42	7.49	9.81
	Cases	41	121	68	22	10	12	275
Electricity and Magnetism	Mean	71.53	67.44	69.36	69.66	71.20	70.50	68.97
	S. D.	9.39	12.31	11.02	10.53	10.26	10.29	11.49
	Cases	31	86	55	19	9	11	205

achievement, as seems reasonable, but if so, the number of sciences studied in high school does not indicate this interest.

A question which immediately arises is the effect of previous courses in university physics on succeeding courses. A canvass of the records shows that all students enrolled in the class in magnetism and electricity had had one or more quarters in mechanics, but there were a great many who had not taken the course in Heat nor the course in optics. The data in Table XCVII are given in order to make comparisons of those who had not taken either Heat or optics, with others of the total group in final ratings in magnetism and electricity. Records for all combinations were not available. There is some evidence showing that students having taken mechanics alone previous to their course in magnetism and electricity do not rate as high, on the average, as those having Heat and optics, or Heat alone, in addition. Whether the course in Heat selects a higher grade of student, or whether the benefits derived from the course carry over into magnetism and electricity is not evident.

TABLE XCVII

MEANS AND STANDARD DEVIATIONS OF FINAL RATINGS IN THE UNIVERSITY COURSE IN MAGNETISM AND ELECTRICITY FOR STUDENTS TAKING DIFFERENT COURSES IN UNIVERSITY PHYSICS

	Courses		
	Mechanics, Heat, Optics	Mechanics, Heat	Mechanics
Mean	71.94	71.83	62.76
S. D.	9.88	11.87	9.70
Cases	47	94	64

CONCLUSIONS

The important conclusions which may be drawn from this study are as follows:

1. No differences in achievement between the members of large and small class groups due to class size are noticeable.
2. New type objective examinations in magnetism and electricity are not found to be any more valid than conventional examinations, using the final quarter mark as a criterion of validity.

3. No valid criteria for pairing students for experimental work in method were found in this study.
4. Students having had high school physics show a tendency on the average to rate higher in the university course in magnetism and electricity at Minnesota, but they are not markedly superior to those not having had the high school course.
5. Students having had high school physics rate significantly higher, on the average, in a test covering the content of high school work in magnetism and electricity than those not having had the high school course.
6. Students in both *high school* and *university* show surprisingly low percentages of correct responses to items covering the respective courses in electricity and magnetism. Either the expectations of instructors are too high, the questions are too difficult, or the students do not measure up to reasonable standards.
7. The number of sciences studied in high school serves as no reliable criterion of achievement in the university courses in physics.
8. Students taking mechanics and Heat or mechanics, Heat, and optics in the University do better, on the average, than those taking mechanics alone, in the course in electricity and magnetism. The students taking Heat alone in addition to mechanics do as well as those taking optics and Heat.

BIBLIOGRAPHY

I. Books

Curtis, Francis D., *Investigations in the Teaching of Science* (Philadelphia: Blakiston's Son & Co., 1926).

Downing, E. R., *Teaching Science in the Schools* (Chicago: The University of Chicago Press, 1925).

Freeman, F. N., *Experimental Education* (Boston: Houghton Mifflin Co., 1916).

Garrett, H. E., *Statistics in Psychology and Education* (New York: Longmans, Green & Co., 1926).

Hudelson, E., *Class Size at the College Level* (Minneapolis: The University of Minnesota Press, 1928).

Klapper, Paul, *College Teaching, Studies of the Methods of Teaching in the College* (Yonkers: World Book Co., 1920).

McCall, W. A., *How to Measure in Education* (New York: The Macmillan Co., 1922).

————, *How to Experiment in Education* (New York: The Macmillan Co., 1923).

MacPhail, Andrew H., *The Intelligence of College Students* (Baltimore: Warwick & York Co., 1924).

Monroe, W. S., *An Introduction to the Theory of Educational Measurements* (New York: Houghton Mifflin Co., 1923).

Pearson, Karl, *Tables for Statisticians & Biometricians* (London: Cambridge University Press, 1914).

Pressey, S. L. and Others, *Research Adventures in University Teaching* (Bloomington: Public School Publishing Co., 1927).

Ruch, G. M., *The Improvement of the Written Examination* (Chicago: Scott, Foresman & Co., 1924).

———— and Stoddard, G. D., *Tests and Measurements in High School Instruction* (Yonkers: World Book Co., 1927).

Rugg, H. O., *Statistical Methods Applied to Education* (New York: Houghton Mifflin Co., 1917).

Rusk, R. R., *Introduction to Experimental Education* (New York: Longmans, Green & Co., 1922).

Russell, Charles, *Classroom Tests* (New York: Ginn & Co., 1926).

Symonds, P. M., *Measurement in Secondary Education* (New York: The Macmillan Co., 1927).

Thorndike, E. L., *Theory of Mental and Social Measurements* (New York: Bureau of Publications, Teachers College, Columbia University, 1919).

Trow, William C., *Scientific Method in Education* (New York: Houghton Mifflin Co., 1925).

Wolf, A., *Essentials of Scientific Method* (New York: The Macmillan Co., 1925).

Wood, Ben D., *Measurement in Higher Education* (Yonkers: World Book Co., 1923).

II. Bulletins, Monographs, and Reports

Barker, F. D., *Determining the Fitness of the Pre-Medical Student,* Bulletin of the Association of American Medical Colleges, 2: 15-20 (January, 1927).

Cabot, Hugh, *The Pre-Medical Course,* Bulletin of the Association of American Medical Colleges, 1: 1-3 (July, 1926).

Capen, S. P., *Pre-Medical Education,* Bulletin of the Association of American Medical Colleges, 1: 4-9 (July, 1926).

Carpenter, W. W., *Certain Phases of the Administration of High School Chemistry,* Teachers' College, Columbia University, Contributions to Education, No. 191, 1925.

Colton, H. S., *The Bearing of High School Science on a College Course in General Zoology,* University of Texas Bulletin, 2:69-72 (July 10, 1915).

Cunningham, H. A., *Individual Laboratory Work versus Lecture Demonstration,* Proceedings of the High School Conference, University of Illinois Bulletin 18, No. 14, 1920.

Gates, Arthur I., *Recitation as a Factor in Memorizing,* New York Archives of Psychology, No. 40, 1917.

Hildebrand, J. H., *Early Training of Scientists,* Bulletin of the American Association of University Professors, 8: 31-32 (May, 1922).

Hurd, A. W., *The Topical versus the Problem Method in the Teaching of High School Physics,* Bulletin of the University of Minnesota, Educational Research Bulletin, 28: 3-9 (January, 1925).

Jones, Harold E., *Experimental Studies of College Teaching,* New York Archives of Psychology, No. 68, 1923.

Mathews, E. J., *The Correlation of High School and College Courses in the Sciences,* University of Texas Bulletin, No. 1838, 1918.

Mayman, J. E., *Teaching Elementary Science in Elementary Schools,* New York Department of Education, Division of Research, Publication 13, 1915.

Monroe, W. S., *The Present Status of Written Examinations and Suggestions for Their Improvement,* University of Illinois Bulletin, Vol. 21, No. 17, Bureau of Educational Research, College of Education (November 26, 1923).

Redfield, A. C., *An Experiment in Teaching Physiology,* Bulletin of the Association of American Medical Colleges, 2: 214-217 (July, 1927).

Seashore, C. E., *Learning and Living in College,* University of Iowa Studies, Volume II, No. 1, 1927.

Spence, Ralph B., *The Improvement of College Marking Systems,* Bureau of Publications, Teachers' College, Columbia University, 1927.

Stebbins, Joel, *Cooperation in Research*, Bulletin of the American Association of University Professors, 8:20-21 (March, 1922).

Whipple, G. M., *Report of Committee on Teaching Experiments*, American Psychological Association, 1913.

Wycoff, John, *Relation of Collegiate to Medical Student Scholarship*, Bulletin of the Association of American Medical Colleges, 2:1-15 (January, 1927).

III. Periodicals

Anderson, J. E. and Spencer, L. T., "The Predictive Value of the Yale Classification Tests," *School and Society*, 24:305-312 (September 4, 1926).

Anibel, Fred G., "Comparative Effectiveness of the Lecture Demonstration and Individual Laboratory Method," *Journal of Educational Research*, 13:355-365 (May, 1926).

Bane, Charles L., "The Lecture vs. the Class-Discussion Method of College Teaching," *School and Society*, 21:300-302 (March 7, 1925).

Bingham, E. C., "Research in Colleges," *Science*, 61:174-176 (February 13, 1925).

Bolenbaugh, L. and Proctor, W. M., "Relation of Subjects Taken in High School to Success in College," *Journal of Educational Research*, 15:87-92 (February, 1927).

Bott, E. A., "The Predictive Value of College Marks in Medical Subjects," *Journal of Educational Research*, 12:214-227 (October, 1925).

Bourne, Randolph, "Experimental Education," *New Republic*, 10:345-347 (April 21, 1917).

Bowers, W. G., "Some Educational Values in Laboratory Work," *Education*, 44:546-555 (May, 1924).

————, "The Advantages of Laboratory Work in the Study of Elementary Chemistry," *Journal of Chemical Education*, 1:205-209 (November, 1924).

———— and Brown, A. F., "Status of Laboratory Work in the High Schools of the Country," *School Science and Mathematics*, 24:815-822 (November, 1924).

Breed, F. S., "Factors Contributing to Success in College Teaching," *Journal of Educational Research*, 16:247-253 (November, 1927).

Carter, T. M., "A Standardized Mental Test vs. an Unstandardized English Test as a Means of Predicting Success in College English," *School and Society*, 26:151-152 (July 30, 1927).

Cleeton, G. W., "The Predictive Value of Certain Measures of Ability in College Freshmen," *Journal of Educational Research*, 15:357-370 (May, 1927).

Colton, H. S., "Segregation of Zoology Students on the Basis of Ability," *School and Society*, 20:254-256 (August, 1924).

COLTON, H. S., "Drawing a Factor in the Training of Students in a Course in General Zoology," *School and Society*, 24:463-464 (October 9, 1926).

————, "Research—A Neglected Function of Elementary Science Courses in the University," *School and Society*, 11:437-438 (April 10, 1920).

CONKLIN, EDWIN, "The Aims in General Biology at Princeton," *School and Society*, 21:429-431 (April 11, 1925).

COOPRIDER, J. L., "Oral versus Written Instructions and Demonstration versus Individual Work in High School Science," *School Science and Mathematics*, 22:838-844 (December, 1922).

————, "Laboratory Methods in High School Science," *School Science and Mathematics*, 23:526-530 (June, 1923).

————, "Teacher versus Student Demonstrations in High School Biology," *School Science and Mathematics*, 26:147-155 (February, 1926).

CRAWFORD, C. C., "Some Experimental Studies of the Results of College Note-Taking," *Journal of Educational Research*, 12:379-386 (December, 1925).

CUNNINGHAM, H. A., "Laboratory Methods in Natural Science Teaching," *School Science and Mathematics*, 24:708-715, 848-851 (October and November, 1924).

CURTIS, F. D., "A Determination of Scientific Attitudes," *Journal of Chemical Education*, 3:920-927 (August, 1926).

DENNY, E. C. and NELSON, M. J., "The Terman and Thurstone Group Tests as Criteria for Predicting College Success," *School and Society*, 26:501-502 (October 15, 1927).

DOWNING, E. R., "A Comparison of the Lecture Demonstration and the Laboratory Methods of Instruction in Science," *School Review*, 33:688-697 (November, 1925).

————, "Relative Value of Laboratory Instruction," *School and Society*, 19:769-770 (June 28, 1924).

DUDDY, E. A. "Report on an Experiment in Teaching Method," *Journal of Political Economy*, 32:582-603 (October, 1924).

EDMONSON, J. B. and MULDER, F. J., "Size of Class as a Factor in University Instruction," *Journal of Educational Research*, 9:1-12 (January, 1924).

ENGLISH, H. B., "The Predictive Value of Intelligence Tests," *School and Society*, 26:783 (December 17, 1927).

ERNST, J. L., "Psychological Tests versus the First Semester's Grades as a Means of Academic Prediction," *School and Society*, 18:419-420 (October 6, 1923).

FOLEY, ARTHUR L., "The College Students' Knowledge of High School Physics," *School Science and Mathematics*, 22:601-612.

FOSTER, H. H., "Lecturing vs. Class Discussion," *School and Society*, 21:386-387 (March 28, 1925).

GARBER, ELINOR, "The Project Method in Teaching Chemistry," *School Science and Mathematics*, 22: 65-73 (1922).

GATES, A. I. and Others, "A Modern Systematic versus an Opportunistic Method of Teaching," *Teachers' College Record*, 27: 679-700 (April, 1926).

GOOD, CARTER V., "An Experimental Study on the Merits of Extensive and Intensive Reading in the Social Sciences," *School and Society*, 23: 758-760 (December 12, 1925).

HAGGERTY, M. E., "The Improvement of College Instruction," *School and Society*, 27: 25-37 (January 14, 1928).

HERRING, J. P., "The Reliability of Accomplishment Differences," *Journal of Educational Psychology*, 15: 530-538 (November, 1924).

HINES, H. C., "Scientific Study of Instruction," *School and Society*, 21: 303-307 (March 14, 1925).

HOLROYD, INA E., "An Analysis of an Experiment in Teaching First Year Mathematics," *School Science and Mathematics*, 22: 114-121 (1922).

HOLTON, E. L., "A Study of Methods of Presenting Subject Matter to Undergraduates in College," *School and Society*, 11: 58-59 (January 10, 1920).

HULL, CLARK L., "The Correlation Coefficient and Its Prognostic Significance," *Journal of Educational Research*, 15: 327-338 (March, 1927).

HUNTER, GEORGE W., "An Experiment in the Use of Three Different Methods in the Class Room," *School Science and Mathematics*, 21: 875-890 (1921).

HURD, A. W., "Subject Matter versus Character Traits," *Journal of Educational Method*, 5: 147-154 (December, 1925).

KELLEY, T. L., "A Simplified Method of Using Scaled Data for Purposes of Testing," *School and Society*, 4: 34-37; (July 1, 1916); 71-75 (July 8, 1916).

KIEBLER, E. W. and WOODY, CLIFFORD, "The Individual Laboratory versus the Demonstration Method of Teaching Physics," *Journal of Educational Research*, 7: 50-58 (January, 1923).

KNOX, W. W., "The Demonstration Method versus the Laboratory Method of Teaching High School Chemistry," *School Review*, 35: 376-386 (May, 1927).

LAIRD, D. A., "Report on Students' Attitudes toward Laboratory Courses," *School and Society*, 19: 589-592 (May 17, 1924).

————, "How the College Student Responds to Different Incentives to Work," *Pedagogical Seminary*, 30: 366-370 (December, 1923).

LANGLIE, T. A., "The Administration of Placement Examinations," *School and Society*, 24: 619-620 (November 13, 1926).

LEWIS, JOHN R., "A Laboratory Examination in First Year College Chemistry," *Journal of Chemical Education*, 4: 1176-82 (September, 1927).

LINDSAY, E. E., "Laboratory Costs in State Institutions of Higher Learning," *School and Society*, 20: 537-542 (October 25, 1924).

McCall, W. A., "What's the Matter with Psychology and Measurement?" *Journal of Educational Research*, 16: 179-183 (October, 1927).

————, "Does It Pay to Measure the Achievement of Pupils?" *Teachers' College Record*, 26: 112-116 (October, 1924).

———— and Johnson, A. E., "A Comparison of Open-Air with Indoor Classes," *Teachers' College Record*, 19: 352-368 (September, 1918).

Mabee, Fred C., "A Test of Achievement in College Chemistry and Results Obtained by Its Use with Both High School and College Classes," *Journal of Chemical Education*, 3: 70-76 (January, 1926).

Mann, C. R., "Cooperative Experiments in Education," *Educational Record*, 6: 39-43 (January, 1925).

Mather, Cleveland, "A New Type of Examination for Medical Students," *Journal of the American Medical Association*, 87: 550-554 (August 21, 1926).

Meyer, A. W., "Laboratory: A Devil's Smithy or a Sanctuary?" *Open Court*, 39: 731-739 (December, 1925).

Miller, H. W., "Segregation on the Basis of Ability," *School and Society*, 26: 84-88 (July 16, 1927), 114-120 (July 23, 1927).

Moore, H. T., "The Attention Value of Lecturing without Notes," *Journal of Educational Psychology*, 10: 467-469 (1919).

Nash, H. B. and Phillips, M. J. W., "A Study of the Relative Value of Three Methods of Teaching High School Chemistry," *Journal of Educational Research*, 15: 371-379 (May, 1927).

Paterson, D. G., "Do New and Old Type Examinations Measure Different Mental Functions?" *School and Society*, 24: 246-248 (August 21, 1926).

Patterson, H., "An Experiment in Supervising College Teaching," *School and Society*, 21: 146-147 (January 31, 1925).

Persinger, C. E., "Experiment in Quantity-Quality Assignment in College Classes," *Historical Outlook*, 17: 293-295 (October, 1926).

Peters, C. J., "An Evaluation of a Reorganization of the Present Core of Subject Matter in High School Physics," *School Science and Mathematics*, 27: 172-182 (1927).

Phillips, T. D., "A Study of Notebook and Laboratory Work as an Effective Aid in Science Teaching," *School Review*, 28: 451-453 (June, 1920).

Pressey, S. L., L. W., and Corbett, Helen, "Research Adventures in University Teaching," *School and Society*, 20: 635-638 (November 15, 1924).

Pyle, W. H., "The Relation of Ability to Achievement," *School and Society*, 22: 406-408 (September 26, 1925).

————, "The I.Q. and the Individual," *School and Society*, 26: 788-790 (December 17, 1927).

Rich, Stephen G., "The Available Tests for Results of Teaching the Sciences," *School Science and Mathematics*, 26: 845-852 (November, 1926).

Riedel, F. A., "What, if Anything, Has Really Been Proved as to the Relative Effectiveness of Demonstration and Laboratory Methods in Science?" *School Science and Mathematics*, 27: 512-519, 620-631 (1927).

SCHEIDEMANN, N. V., "A Comparison of Two Methods of College Instruction," *School and Society*, 25:672-674 (January 4, 1927).

SCHUTTE, T. H., "Is There Value in the Final Examination?" *Journal of Educational Research*, 12:204-213 (October, 1925).

SEASHORE, C. E., "College Placement Examinations," *School and Society*, 20:575-578 (November 8, 1924).

SPENCE, R. B., "Factors Related to College Achievement," *Teachers' College Record*, 29:504-514 (March, 1928).

STODDARD, G. D., "Iowa Placement Examination," *School and Society*, 24:212-216 (August 14, 1926).

SYMONDS, P. M., "Needed Research in the Field of Measurement in Secondary Education," *Journal of Educational Research*, 16:119-126 (September, 1927).

THORNDIKE, E. L., "Measurement in Education," *Teachers' College Record*, 22:371-379 (November, 1921).

———— and Others, "Standard Tests and Their Use," *Teachers' College Record*, 26:93-116 (October, 1924).

TOOPS, H. A., "The Use of Intelligence Tests by Universities," *School and Society*, 24:87-88 (July 17, 1926).

VAUGHN, JAMES, "The Comparative Value of Certain Measures for Predicting Grades in College Physics," *School and Society*, 18:28-30 (July 7, 1923).

WAKEMAN, G., "Education and the College Professor," *School and Society*, 26:812-814 (December 24, 1927).

WAPLES, DOUGLAS, "A Technique for Investigations in Class Room Methods," *Journal of Educational Research*, 11:254-268 (April, 1925).

WILEY, W. H., "An Experimental Study of Methods in Teaching High School Chemistry," *Journal of Educational Psychology*, 9:181-198 (April, 1920).

WINCH, W. H., "Can Observation Be Trained in Children?" *Journal of Educational Research*, 15:229-238 (April, 1927); 314-326 (May 16, 1927).

WOOD, BEN D., "Relation of College Curricula to Educational and Vocational Guidance," *School Life*, 10:127-129.

————, "Studies of Achievement Tests," *Journal of Educational Psychology*, 17:1-22 (January, 1926); 17:125-139 (February, 1926); 17:263-269 (April, 1926).

WOODY, CLIFFORD, "A Survey of Educational Research in 1923," *Journal of Educational Research*, 9:357-381 (May, 1924).

YOUNG, J. B., "How Emotional Traits Predispose to College Failure," *Journal of Educational Psychology*, 18:631-637 (1927).

INDEX

Abilities, developed by laboratory work, 11, 128-129
Achievement, group, in anatomy, 32-35; in physiology, 59-61, 68-69; in mechanics, 89-97, 136-137; in heat, 140-142; in magnetism and electricity, 170-172, 178-181
Achievement, individual, in anatomy, 41-47; in physiology, 61, 68; in mechanics, 119-120, 122; in heat, 147; in magnetism and electricity, 165
Achievement, intelligence and, 101-102; determinants of, 161-184; interest and, 182-183
Achievement, prediction of, in anatomy, 50; in physiology, 62-63; in mechanics, 102-103, 105, 115; in heat, 147-148; in magnetism and electricity, 162, 169
Aims and objectives, the determination of, 2, 5-6; of the laboratory, 9-12, 128-130; in mechanics, 133-134; in the light of achievement, 179
Anatomy, laboratory work in, 25; experiment in, 25-54; plan of instruction in, 31; achievement in, 32-35, 39-47; 1925-26, 47-49; discussion of experiment in, 50-54
Anibel, 17
Architectural students, 102, 116-117, 119-120, 123
Aristotle, 8
Authority and experimentation, 8

Bane, C. L., 14
Bing, Ada, 16-17
Biology, 19
Borgersrode, F. von, 16, 142
Buchta, J. W., 89, 90

Carpenter, W. W., 18
Character traits, 18, 36
Chemistry, 15, 19-20
Chicago, University of, 10
Class size, in heat, 139-148; in magnetism and electricity, 161-166
College groups, achievement in mechanics, 100
College instruction, 6-7
Colton, H. S., 13
Committee on Educational Research, 4; program of investigation, 5; sub-committee on teaching of science, 7
Conclusions and recommendations, 151-157
Cooperation in research, 3-4
Cooprider, J. L., 17
Correlations, anatomy and other ratings, 49; between tests in anatomy, 50; physiology and other ratings, 62; pre-test in physics and intelligence, 78; raw scores and weighted scores, 79; for reliability of pre-tests in mechanics, 83-84; for reliability and validity of achievement tests in mechanics, 96-97; mechanics and other ratings, 103, 105-114; between conventional and new type test, 141; between mechanics and heat, 147; for Iowa Placement Test, 148; magnetism and electricity and other ratings, 162, 169; for reliability of conventional tests, 167.
Correlation charts, 106-114.
Cost of instruction, 2, 5, 26, 55
Crawford, C. C., 16
Criteria for judgment, 153-154
Critical ratings, 105-114
Cunningham, 17
Curriculum in science, 5
Curtis, F. D., 18

Downing, E. R., 17
Duddy, E. A., 15
Duell, H. W., 20

Edmonson and Mulder, 16
Education, 14
Educational psychology, 14
Electricity and magnetism, class size in, 161-166; criteria of success in, 166-169; high school physics and achievement in, 170; preliminary test in, 171; achievement in college, 180-181; interest and achievement in, 181-182;

previous courses in physics as a prognosis of success in, 183; some conclusions from the study in, 183-184
Equation of experimental groups, in high school physics, 17-18; in high school chemistry, 20; in human anatomy, 27-30, 36-37; in physiology, 56-58, 65-67; discussion of, 102; in mechanics, 116, 140; in magnetism and electricity, 162-163
Equivalence of forms of tests, 93
Erikson, H. A., 90
Experimental attitude, 1, 7
Experimentation, need for, 8; techniques for, 154-155. *See* Research, educational

Faculty opinion, 9-10, 51-53
Foley, A. L., 85
Freeman, E. M., 7

Galileo, 8
Garber, E., 16
Good, C. V., 15

Heat, class size in, 139-148; group comparisons in, 142-146; prognosis in, 146
High school students, achievement in physics, 84-87
Holroyd, I. E., 17
Hudelson, E., 16, 142
Hull, C. L., 105
Hunter, G. W., 16
Hurd, A. W., 6, 16-18, 21, 83, 162
Hurd tests, 75-76, 81, 85, 171

Instruction, features of science, 11, 129; methods of, *see* Methods of instruction
Investigations in science, 18
Iowa, University of, 14
Iowa physics test, 19
Iowa placement test, 15, 139

Jackson, C. M., 26, 31
Johnson, P. O., 19

Kiebler and Woody, 17
Knox, W. W., 19

Laboratory, aims; *see* Aims and objectives
Laboratory test, 90-92, 120
Lyon, E. P., 55

McCall, W. A., 16, 18-19, 81-83, 162
M scores, 83, 104, 123, 168-169
Magnetism and electricity, *see* Electricity and magnetism
Mayman, J. E., 16
Measurements of achievement, as a problem, 5-6, 8; in college general zoology, 13; in education, 14; in chemistry, 15, 20; in anatomy, 32-35, 39-47; in physiology, 59-61; in physics, 81-84; in mechanics, 89-98; in heat, 140-141; in magnetism and electricity, 166-169
Mechanics, plan of experiment in, 102, 116; critical ratings for, 105-114; group comparisons, 116-125; group summaries, 121-122; classes 1923-27 in, 123-125; effect of laboratory work in, 123, 125; student opinion in, 126-132; distribution of letter ratings in, 132; discussion of student opinion in, 132-135
Medicine, subsequent course records in, 53-54
Mental ability, different college groups, 77-78; and high school physics, 86-87
Method, laboratory, 1, 8, 17, 19, 25, 55, 64, 75; lecture-conference, 14; lecture, 14; class-discussion, 14; individualized, 14, 18-21; demonstration, 11, 18-19; group-of-two, 18; control, 18; teacher-assigned, 18-19; pupil-assigned, 18-19; reorganized, 19; textbook, 19, 21; topical, 21; problem, 21
Methods of instruction, investigation of, 5, 7; value of in science, 11; former experiments in, 13-21; in anatomy, 25-26; in physiology, 55, 64; in mechanics, 75, 102, 125, 129, 134
Miller, W. S., 27
Minnesota, University of, 9, 12, 15-16
Mulder and Edmonson, 16

Nash and Phillips, 20
Noll, V. H., 15
Norms on physics tests, 75, 85

Objectives, *see* Aims and objectives
Objective tests versus conventional tests, 98, 141, 167-168
Otis, A. S., 141; correlation chart, 49, 79

Pairing of groups, *see* Equation of groups
Pennsylvania, University of, 13
Persinger, C. E., 15
Peters, C. J., 19
Phillips, 17
Phillips and Nash, 20
Physics, investigations in secondary school, 16-18, 20-21, 83; tests in secondary school, 19, 75-76, 81, 85, 171; high school student achievement in, 84-86; mental ability and high school, 86-87; effect of high school in university, 99, 170; knowledge of high school, 173-179; knowledge of university, 181, 183. *See* Electricity and magnetism, Mechanics, Heat
Physiology, laboratory vs. library, 55; experiments in, 55-71; plan of experiment, 55-56, 64; achievement in, 59-61, 67-69; reducing laboratory time, 64-71; conclusions for experiment in, 70; subsequent data for, 70-71
Prediction of achievement, *see* Achievement, prediction of
Preliminary tests, in mechanics, 75-76, 81-82; in magnetism and electricity, 171
Problems of college education, 5
Prognostic tests, development of, 136-138

Questionnaire, 9-12, 126-132

Recommendations, and conclusions, 151-157
Reliability, of differences, in anatomy, 29-31, 37, 40-41; in physiology, 57-58, 60, 62, 66, 69; in mechanics, 86-87, 93, 100-101, 119, 124; in heat, 146; in magnetism and electricity, 165, 170, 172
Reliability, of tests, in anatomy, 51; in mechanics, 83-84, 96, 97; in magnetism and electricity, 162, 167
Research, educational, 4, 12-21; methods of, 6, 13-21, 153-155
Riedell, F. A., 17
Rotation plan, 18-20
Rugg, H. O., 79

Scaling of tests, preliminary in mechanics, 79-83; by Rugg technique, 79-81; total scores, 80-83, 90-92, 168-169; value of, 78-79, 81
Scheidemann, N. V., 14
Scientific method, 1-3
Scoring, manner of, 94-96
Scott, F. H., 55
Secondary school research, *see* Studies
S. D. percentiles, computation of, 80
Significance of tests, 97
Spearman-Brown formula, 84
Student opinion, in science, 10-12; in anatomy, 53; in mechanics, 126-138; preference for tests, 99, 130-131
Studies, at the college level, 13-16; at the secondary school level, 16-21

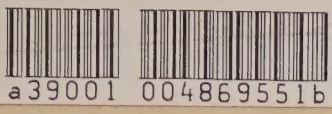
Techniques for experimentation, 154-155
Tests, in high school physics, 75-78, 81, 84, 162, 171-172, 179; in college physics, 89-101, 136-138, 140-141, 165-167, 169
Thorndike, E. L., 29
Twin City vs. other high schools, 87-88

Validity of tests, in anatomy, 51; in mechanics, 83-84, 96-98; in heat, 140-141; in magnetism and electricity, 167-168

Wiley, 17
Woody and Kiebler, 17

Zoology, 13

Q181. H97 1928


a39001 004869551b

Q
181
H9
192

68-1